The nucleation and growth of meta-aminobenzoic acid: a density functional theory and molecular dynamics study.

Etienne Gaines

Submitted in partial fulfillment

of the requirements of the Degree of

Doctor of Philosophy

School of Biological and Chemical Sciences

Queen Mary University of London

09.08.2018

"So sleeps the pride of former days."

Thomas Moore

Statement of originality

I, Etienne Gaines, confirm that the work presented in this thesis is my own. The use of all materials from other sources has been properly and fully acknowledged in the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Abstract

Controlling crystal polymorphism, the ability of a molecule to crystallise in different solid forms, is one of the grand, ongoing challenges in materials science. In the pharmaceutical industry particularly, where up to half of the active pharmaceutical active ingredients exhibit polymorphic behaviour, it is of paramount importance to rationalise the impact of experimental conditions, such as the nature of the solvent, on the obtainment of a specific crystal form. As strategies for the selection of polymorphs is still, by and large, based on a trial-and-error approach, it is necessary to acquire a fundamental understanding of the factors controlling the formation of a specific solid-state structure during crystallisation from solution.

During this doctoral research project, we have conducted a computer simulation study of the early stages of crystallisation of meta-aminobenzoic acid, an important model system in the investigation of polymorphic phenomena. This molecule can in fact form five different polymorphic forms whose selective crystallisation from solution chiefly depends on the nature of the solvent.

Molecular models and computational chemistry methods, based on density functional theory and molecular dynamics, have been developed and applied to quantify the processes surrounding the crystallisation of meta-

aminobenzoic acid: solvent-solute separation, solute aggregation and surface reactivity. The aim was to identify what controls, at the molecular level, the polymorphic selection process during crystallisation from solution of this important active pharmaceutical ingredient.

The results show that the solvent play a significant role during the key stages of meta-aminobenzoic acid crystallisation by controlling both the kinetics and thermodynamics of solute desolvation, formation of prenucleation clusters and surface reactivity. This work represents a paradigm of the role of molecular processes during the early stages of nucleation in affecting polymorph selection during crystallisation from solution.

Acknowledgements

I would like to thank Dr. Devis Di Tommaso for his constant help and support during these three years in Queen Mary. I would also like the thank the other members of my panel, Dr. Isaac Abrahams and Dr. Gregory Chass, for their countless advices on the organisation and the orientation of my research or simply my writing skills.

Last but not least, I need not to forget all those along whom I worked for this period, Post-docs and PhD students alike, I thank you for bearing with me.

This research utilised Queen Mary's MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC grand EP/K000128/1. Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work used the ARCHER UK National Supercomputing Service.

Contents

1 Introduction **20**

1.1 Polymorphism 20

1.2 Crystallisation 22

1.2.1 Desolvation 23

1.2.2 Nucleation 25

1.2.3 Growth 31

1.3 Solvent effect on polymorph control 33

1.4 Previous computational studies 35

1.5 Meta-aminobenzoic 38

1.6 Aim of the project 40

1.7 Thesis outline 41

2 Theoretical methods **43**

2.1 Density functional theory 43

2.1.1 Kohn-Sham approach 46

2.1.2 Jacob's ladder 48

2.1.3 Finding the minimum 49

2.2 Forcefield 51

2.3 Molecular dynamics 53

2.3.1 The equation of motion 53
2.3.2 Statistical mechanics 55
2.3.3 Modelling the physical system 56
2.3.4 Integration algorithms 56
2.3.5 Periodic boundary conditions 59
2.4 Computational procedures 60
2.4.1 Simulation software 60
2.4.2 In-house codes 63

3 Solvation of meta-aminobenzoic acid in aqueous and organosulfur solutions 66

3.1 Introduction 67
3.2 Methods 69
3.2.1 Ab initio molecular dynamics simulation 69
3.2.2 Static density functional theory calculations 70
3.3 Hydration of meta-aminobenzoic acid 72
3.4 Equilibrium of zwitterionic non-zwitterionic forms in DMSO 81
3.5 Quantification of the strength of mABA-solvent interaction . 84
3.6 Conclusions 86

4 The role of solvent in the self-assembly of m-aminobenzoic acid 88

4.1 Introduction 89
4.2 Methods 91
4.2.1 Density functional theory calculations 91
4.2.2 Classical molecular dynamics 95
4.3 Role of solvent in the energetics of cluster formation 97

4.3.1 Dimers 98
4.3.2 Trimers 102
4.3.3 Tetramers 103
4.4 Role of solvent in the kinetics of aggregation 105
4.5 Free energy profiles of dimerisation 108
4.6 Conclusions 110

5 Aggregation of Meta-aminobenzoic Acid in Water 112
5.1 Introduction 113
5.2 Computational methods 115
5.2.1 Density functional theory calculations 115
5.2.2 Molecular dynamics simulations 118
5.3 Dimerisation of meta-aminobenzoic acid 120
5.4 Molecular aggregation in mixed mABA/mABA$^{\pm}$ aqueous solutions 123
5.5 Conclusions 128

6 Surface stability and reactivity of form-II of meta-aminobenzoic acid 130
6.1 Introduction 131
6.2 Computational methods 134
6.2.1 Surface construction 134
6.2.2 Density functional theory calculations 134
6.2.3 *Ab initio* molecular dynamics 135
6.2.4 Classical molecular dynamics 136
6.3 Stability of mABA surfaces 139
6.4 Adsorption of building units on stable mABA surfaces . . . 143

6.5 Crystal-solvent interactions 145
6.6 Conclusions 148

7 Conclusions **150**

A Appendix **156**

A.1 Tables 156
A.2 GAFF parameters 157
A.3 Codes in Fortran 173
A.3.1 COMFORT 173
A.3.2 configuration_selector and distance_test subroutine . . 181

List of Figures

1.1 Schematic relationship between the difficulty of nucleation of salicylic acid as determined by calorimetric measurements and Raman spectroscopy and the strength of the solute-solvent interaction as computed by density functional theory by Khamar *et al.*[32] 24

1.2 Free energy of the system as a function of crystal units. (a) Classical pathway. (b) and (c) Prenucleation pathways. PNCs of defined size increase the barrier to nucleation (b), whilst metastable disperse PNCs lower it, acting as building blocks to crystal growth (c). 27

1.3 Nucleation pathways, adapted from Davey *et al.*[39] 28

1.4 A schematic representation of classical growth theory from Teng[42] 31

1.5 Tetrolic acid molecules in different solvents (left), in pre-nucleation aggregates (middle) and in their crystal form (right) from Chen *et al.*[76] 34

1.6 Renucleation pathways of molecules in solution starting from different polymorphs from Nordström *et al.*[83] 35

1.7 mABA molecule in its nonionic (left) and zwitterionic (right) forms. 38

1.8 The five known polymorphs of mABA with zwitterionic (left) and non-zwitterionic (right) forms adapted from Svard *et al.*[99] and Williams *et al.*[95] 40

2.1 Representation of *Jacob's ladder* density functional approximations for the exchange correlation term by Perdew *et al.*[103] 48

2.2 From left to right, visualisation of the force-field parameters for bonds, angles, torsions, electrostatic and Van der Waals energies. 53

3.1 Schematic picture of the two tautomeric forms of meta-aminobenzoic acid: (a) nonionic mABA; (b) zwitterionic $mABA^{\pm}$. The oxygen and nitrogen atoms of mABA or $mABA^{\pm}$ are denoted by O_m and N_m, the hydrogen of amino group are denoted by H_a, the hydrogen atoms of carboxylic group are denoted by H_c. 68

3.2 Time evolution of the X_m-H (X = N, O) distances during the AIMD simulation of the nonionic (mABA) and the zwitterionic ($mABA^{\pm}$) forms of meta-aminobenzoic acid in water: (a) intramolecular (N_m-H) and intermolecular (N_m...H) distances of the mABA molecule; (b) intramolecular (O_m-H) and intermolecular (O_m...H) distances of the mABA molecule; (c) intramolecular (N_m-H) and intermolecular (N_m...H) distances of the $mABA^{\pm}$ molecule; (d) intramolecular (O_m-H) and intermolecular (O_m...H) distances of the $mABA^{\pm}$ molecule. 74

3.3 The radial distribution functions, g(r) (continuous lines), and running coordination numbers, n(r) (dashed lines), of mABA and $mABA^{\pm}$ with water obtained from AIMD simulations: (a) O_m-H_w RDFs (O_m = oxygen atoms of meta-aminobenzoic acid; H_w = hydrogen atoms of water); (b) N_m-H_w RDFs (N_m = nitrogen atoms of meta-aminobenzoic acid; H_w = oxygen atoms of water). The integration number has been computed as $n(r) = \int 4\pi N/V \int_0^r g(r)dr$, where N is the number of hydrogen atoms and V is the volume of the simulation cell. 76

3.4 The radial distribution functions, g(r) (continuous lines), and running coordination numbers, n(r) (dashed lines), of mABA and mABA$^{\pm}$ with water obtained from AIMD simulations: (a) $H_c - O_w$ RDFs (O_c = oxygen atoms of the carboxylic group of mABA; H_w = hydrogen atoms of water); (b) $H_a - O_w$ RDFs N_m = nitrogen atoms of the amino group of mABA and mABA$^{\pm}$; O_w = oxygen atoms of water). The integration number has been computed as $n(r) = \int 4\pi N/V \int_0^r g(r)dr$, where N is the number of hydrogen atoms and V is the volume of the simulation cell. 78

3.5 (a) Probability distribution of the coordination number in the hydration shell of mABA, the mABA-H_2O radial distribution function of the center-of-masses of mABA and water (inset), and the optimised structure of mABA with its hydration shell. (b) Probability distribution of the coordination number in the hydration shell of mABA$^{\pm}$, the mABA$^{\pm}$-H_2O radial distribution function of the center-of-masses of mABA$^{\pm}$ and water, and the optimised structure of mABA$^{\pm}$ with its hydration shell. 80

3.6 Optimised gas phase structures of the most stable solvated mABA$^{\pm}$(DMSO)$_n$ (n = 1-3) clusters as computed at the M06-2X/6-31+G(d,p). 82

4.1 Schematic representation of m-aminobenzoic acid clusters, (mABA)$_n$ (n = 2-4), located in this study. 100

4.2 Fraction of the stoichiometric mABA concentration present as monomers and dimers in DMSO. Curve obtained by applying eqn 4.7 with K_D = 304.9 L.mol^{-1}. 102
4.3 Correspondence between the most stable meta-aminobenzoic acid tetramer, $(mABA)_4$, in solution and the synthon found in the (1 x 3 x 1) unit cell of form II of mABA. 104
4.4 Last configurations of MD simulations of 0.3 mol.L^{-1} and 1.1 mol.L^{-1}. mABA solutions in DMSO (200 ns) and water (30 ns). For clarity, solvent molecules were removed. 106
4.5 Time evolution of the number of pairs between mABA molecules computed during the last 10 ns of the MD simulations of (a) 0.3 mol.L^{-1} and (b) 1.1 mol.L^{-1} solutions. 107
4.6 Free energy profiles for the meta-aminobenzoic acid dimerisation in water and DMSO. The order parameter (OP) used to study this reaction was defined by averaging two distances between the hydroxyl hydrogen and carboxyl oxygen of a pair of mABA molecules. A, B, C and D represent the structure of the microstates along the OP trajectory determined using the *METAGUI* computational tool.[155] 108
4.7 Free energy profile for the mABA dimerisation reaction in water and DMSO obtained by using metadynamics. The order parameter (OP) used to study the dimerisation reaction was the carbon (C=O) carbon (C=O) distance of two mABA molecules. 109

5.1 Crystal structure of the polymorphs of meta-aminobenzoic acid denoted II, III and IV: (a) (1x3x1) unit cell of Form II (neutral); (b) (1x2x2) unit cell of Form III (zwitterionic); (c) (2x2x1) unit cell of Form IV (zwitterionic).[95] 114

5.2 Convergence of the volume of the simulation boxes containing mixed mABA-mABA$^{\pm}$ solutions. 120

5.3 Optimised structures of most stable meta-aminobenzoic acid dimers in water: (a) nonionic $(mABA)_2$ dimer (in parenthesis value obtained using the gas-phase optimised geometries of $(mABA)_2$ and mABA); (b) nonionic-zwitterionic (mABA)(mABA$^{\pm}$) dimer; (c) zwitterionic $(mABA^{\pm})_2$ dimer. Beneath the structure is reported free energy of dimer formation in water. . . 122

5.4 Configuration at 200 ns of mixed mABA-mABA$^{\pm}$ aqueous solutions: (a) 0.04 mol.L^{-1} aqueous solution; (b) 0.09 mol.L^{-1} aqueous solution; (c) 0.16 mol.L^{-1} aqueous solution; (d) 0.31 mol.L^{-1} aqueous solution. Water molecules have been removed. The grey outlines represent the cubic simulation box. 124

5.5 Time evolution of the number of pairs between meta-aminobenzoic acid molecules in mixed mABA-mABA$^{\pm}$ aqueous solutions computed during the last 40 ns of the MD simulations. . . . 125

5.6 Time evolution of the number of pairs between meta-aminobenzoic acid molecules in mixed mABA-mABA$^{\pm}$ aqueous solutions computed during the last 40 ns of the MD simulations: (a) 0.16 mol.L^{-1}; (b) 0.31 mol.L^{-1}. 125

5.7 Three-body representations (A-B-C) and (A*-B*-C*) of the nonionic, mABA, and zwitterionic, $mABA^{\pm}$ forms of meta-aminobenzoic acid: A and A* are the center-of-masses (COMs) of the -NH_2 and -NH_3^+ groups, B and B* are the COMs of the benzine (C_6H_4) group, and C and C* are the COMs of the -COOH and -COO^- groups. 126

5.8 Intermolecular distances between the amino (NH_2 and NH_3^+), carboxylic (COOH and COO^-), and benzine (C_6H_4) groups in the most thermodynamically stable $(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$ dimers in water. 127

6.1 Molecular models of the unstable (100), (top) and (001), (bottom) surfaces of form II of mABA for different number of layers. 140

6.2 Evolution of the surface energies (in $J.m^{-2}$) with the number of layers in contact with a fixed 20 Å vacuum slab. Calculations conducted using the CP2K code with the PBE-D3 functional, the DZVP basis set and a PW cutoff of 1000 Ry. (001) and (00-1) are probably equivalent. 141

6.3 Simplified representation of the dipole moments in the unstable (100) surface (left) and stable (001) surface (right). . . 142

6.4 Arrangements of molecular units of mABA (monomer, dimer and tetramer) on top of the (001) surface of mABA form II. 145

6.5 Final snapshots of molecular dynamics simulations for stable surfaces (001), *left* and (010), *right*, with both water, *top* and DMSO, *bottom*, at the solute-solvent interface. 146

List of Tables

3.1 Positions (r_{max}^{X-H} and r_{min}^{X-H} in Å) and amplitudes (g_{max}^{X-H} and g_{min}^{X-H}) of the maxima and minima of the first peak of the $X_m - H_w$ (X = O_m, N_m) RDFs, and first hydration shell numbers (n_w) obtained from the AIMD simulations of mABA and mABA$^{\pm}$ in water. 77

3.2 Positions (r_{max}^{H-O} and r_{min}^{H-O} in Å) and amplitudes (g_{max}^{H-O} and g_{min}^{H-O}) of the maxima and minima of the first peak of the $H_a - O_w$ and $H_c - O_w$ RDFs, and first hydration shell numbers (n_w) obtained from the AIMD simulations of mABA and mABA$^{\pm}$ in water. 79

3.3 Solvation free energies in DMSO of the zwitterionic (mABA$^{\pm}$) form of m-aminobenzoic acid obtained using the cluster-continuum approach. Gas-phase free energy contributions computed at the M06-2X/6-31+G(d,p) level and solvation free energies computed at the SMD/M06-2X/6-31+G(d,p) level. Values in kcal.mol^{-1} 82

3.4 Energetics of the zwitterionic equilibrium mABA $\rightleftharpoons$ mABA$^{\pm}$. Gas-phase energies ($\Delta E_{e,gas}$) and standard state (1 atm) gas-phase free energies computed with different quantum chemistry methods. Geometries, frequencies and solvation energies computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal.mol^{-1} 84

3.5 Solute-solvent binding energy (ΔE_{clust}), free energy of solutesolvent clustering (ΔG_{clust}) and successive solvent binding energy (ΔE_{inc}) and solvation free energies of m-aminobenzoic acid in water and in DMSO. Geometries, frequencies and solvation energies computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal.mol^{-1} 85

4.1 Details of classical molecular dynamics simulation of mABA solutions 97

4.2 Energetics of formation of meta-aminobenzoic acid clusters in the gas-phase and solution as computed at the M06-2X/6-31+G(d,p) level of theory. Values obtained from the Boltzmann average of the energies or free energies of the low-lying $(mABA)_n$ isomers. Values in kcal.mol^{-1} 98

4.3 Energetics of formation of m-aminobenzoic acid dimers in the gas-phase and solution as computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal.mol^{-1}. 99

5.1 Details of molecular dynamics simulation. 120

5.2 Energetics of dimerisation of meta-aminobenzoic acid: $\Delta E_{e,gas}$ is the gas phase interaction energy; ΔG^{0}_{ass} is the standard state (1 atm) gas-phase association free energy at 298 K; ΔG^{*}_{ass} is the standard state (1 mol.L^{-1}) free energy of reaction in the liquid-phase. Calculations conducted at the M06-2X/6-31+G(d,p) level of theory using the SMD solvation model. Values obtained from the Boltzmann average of the energies, or free energies, of the isomers of nonionic $(mABA)_2$, zwitterionic $(mABA^{\pm})_2$, and mixed $(mABA)(mABA^{\pm})$ dimers. Values in kcal.mol^{-1}. 121

5.3 Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.08 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage. 128

6.1 Binding energies (and binding energies per mABA unit) associated with the adsorption of monomer, dimer and tetramer on the (001) and (010) surfaces of form II of mABA. Values in kcal.mol^{-1}. 144

6.2 Number of exchanges (N_{ex}) of the solvent molecules at the (001) and (010) surfaces of form II of mABA. An exchange is considered to be real if the new position outside or inside the coordination shell of mABA lasts for more 0.5 ps. Values obtained from the analysis of the last 10 ns of every simulation period. Values normalised with respect to the number of mABA at the interface. 147

6.3 Number of exchanges (N_{ex}) of the DMSO molecules at the (001) and (010) surfaces of form II of mABA. An exchange is considered to be real if the new position outside or inside the coordination shell of mABA lasts for more than τ. Values obtained from the analysis of the last 10 ns of every simulation period. Values normalised with respect to the number of mABA at the interface. 148

A.1 Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.04 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage. 156

A.2 Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.16 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage. 157

A.3 Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.31 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage. 157

Chapter 1

Introduction

1.1 Polymorphism

The term polymorphism, from ancient greek *poly* (many) and *morphism* (form), was supposedly used for the first time in crystallographic literature by Mitscherlich in 1822 when he recognised different crystal structures in arsenate and phosphate.[1] Polymorphism defines the ability of a given chemical compound to crystallise in different solid structures.[2] Substances that can exist in several amorphous forms are instead referred as molecules exhibiting *polyamorphism*.[3]

In organic materials science, the first observation of polymorphism is attributed to Wöhler and von Liebig, in 1832, when they examined the cooling of a benzamide solution to discover two different forms: silky needles followed by rhombic crystals.[4] Later studies on benzamide identified an additional polymorph.[5] New polymorphs are not always easy to discover and additional forms might stay unidentified for a long period of time before being observed and characterised. A good example is maleic acid, which

was thought to exist in only one form for more than a *century.*[6] Bodewig reported crystallographic data around 1881 before subsequent X-ray analysis in 1925, 1939[7], 1952[8] and 1974[9], all of which identified the same crystal form. Having become a bulk product in the chemical industry, produced in huge quantities and extensively used in the pharmaceutical industry, the discovery of a second form of maleic acid was a challenging surprise.[6,10] As stated by Professor Sally Price during the British Association of Crystal Growth annual lecture : "We could argue that there are two main reasons for which new polymorphs are sometimes not intuitively found: either good experiment has not been done or the right crystallisation experiment is not available just yet".

Three different types of polymorphism exist: *packing polymorphism* (difference in crystal packing), *conformational polymorphism* (different conformers of the same molecule) and *solvomorphism*[11] (or pseudopolymorphism, mainly obtained from the introduction of solvent molecules in crystalline phases).

In a recent study, partly based on compounds from the Cambridge Structural Database (CSD), it was found that at least 50% of the screened systems exhibited polymorphism.[12] Similar statistical trends were found for co-crystals and salts and, to a lesser extent, for solvates and hydrates. The most prolific polymorphic system is *5-methyl-2-[(2-nitrophenyl)-amino]-3-thiophenecarbonitrile* that has no less than ten reported polymorphs including seven well defined crystal structures.[13] Only eight molecules possess five known polymorphs (including the molecule meta amino-benzoic acid) and thirty-four have four polymorphic structures. Given the size of the CSD, systems exhibiting abundant polymorphism (four or more crystal struc-

tures) are extremely rare and they can be used as models system to better understand what controls the selection of particular polymorphic form during crystallisation from solution.

Polymorph selection is particularly important in the development of active pharmaceutical ingredients (API), as a drug receive regulatory approval only for on single crystal form. In fact, multiple crystalline forms of an API have different stability, bio-availability, uptake in the bloodstream, activity or simply different performance in wider applications. Therefore, overlooking the *problem* of polymorphism is not a possibility and polymorph control during the crystallisation of an API is an ongoing challenge for the pharmaceutical industry. Apart from the necessity to solve this industrial challenge, polymorphic research is also driven from the scientific curiosity of what controls the formation of crystals. When embarking in the study of polymorphism, it is always nice to keep in mind the words of Walter McCrone: "*every compound has different polymorphic forms, and that, in general, the number of forms known for a given compound is proportional to the time and money spent in research on that compound.*" [14]

1.2 Crystallisation

The study and manipulation of crystal forms for academic and artistic purposes can be traced back all the way to the Bronze Age.[15] The practical consequences of crystal nucleation in science and technology are countless and can also be found in everyday life. Climate change is greatly impacted by the crystallisation of ice,[16,17] with how blankets of frozen water insulate

the ground and the oceans for example, while the oil industry is affected by phenomena leading to the crystallisation of hydrocarbon clathrates, that typically form with small molecules of methane or carbon dioxide in contact water at ambient temperature.[18,19] Crystallisation can also happen spontaneously in living beings, whether it's a natural beauty, like the nautilus shell[20] or curse, with the formation of amyloid fibrils, a plausible cause for Alzheimer's disease.[21,22]

The process of crystallisation from solution can be considered to occur through the sequence of the following steps: *desolvation* of the solute molecules; *nucleation* of the solid phase; *growth* of the crystal nuclei to larger sizes by incorporation of material from the solution.

1.2.1 Desolvation

The process of desolvation of the building units (molecules or ions) of a macroscopic crystal represent the first step of the process of crystal growth from solution. In the case of crystals of ionic type A_xB_y, it is generally accepted that cation dehydration is the rate-limiting step to crystal growth and governs the kinetics of nucleation[23] and of reactions at the mineral-water interfaces such as adsorption,[24] crystal growth and dissolution.[25–28] For example, the slow rate of formation of magnesite $MgCO_3$ from ambient solution environments, which represents one of the most intriguing and long-standing research questions in geochemistry,[29] has long been ascribed to the slow kinetics of dehydration of the magnesium ion, Mg^{2+}, which is strongly bound to six water molecules in the first hydration shell.[30]

Recent work has also highlighted the important role of solute-solvent in-

teraction in the kinetics of formation of organic crystals. For salicylic acid, induction time analysis showed that the crystal nucleation in different solvents becomes increasingly more difficult in the following order: chloroform, ethyl acetate, acetonitrile, acetone, methanol and acetic acid.[31] By means of vibration spectroscopy, calorimetric measurements and density functional theory (DFT) calculations, Khamar *et al.*[32] concluded that the influence of the solvent on crystal nucleation is related to the process of desolvation of the salicylic acid molecules: the stronger the interaction between the salicylic acid and the solvent molecules that are part of its first solvation shell, the slower is the crystal nucleation in solution (Figure 1.1).

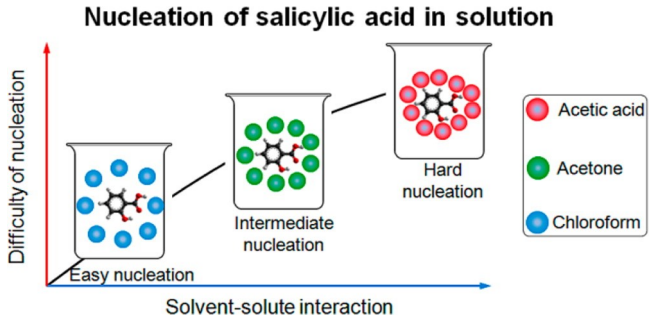

Figure 1.1: Schematic relationship between the difficulty of nucleation of salicylic acid as determined by calorimetric measurements and Raman spectroscopy and the strength of the solute-solvent interaction as computed by density functional theory by Khamar *et al.*[32]

A similar solvent-dependent behavior was observed for the crystal nucleation of risperidone, an API mainly used in the treatment of schizophrenia and bipolar disorder. Here, DFT calculations of the solvent-solute binding interactions linearly correlate with induction time crystallisation experiments of risperidone in different solvation environments.[33]

Consequently, a fundamental description of the structure of the solvent around the solute molecule as well as the energetics of solute-solvent inter-

action can be of fundamental importance to understand, characterise and, in part, predict the propensity of nucleation of an organic crystal from solution.

1.2.2 Nucleation

Crystal nucleation occurs via the formation of a new phase within a large volume of the old phase, in which a small number of atoms, ions or molecules are arranged in a pattern characteristic of the new solid.[34–36] Because of the metastability of the old phase, the transformation of the system has a free energy barrier that has to be overcome.[37] During the formation of the new phase, a special attention has to be drawn upon surface molecules of the emerging phase, as they are less well bound to one another compared to those in the bulk (their contribution to the global free energy of the system is greater). The consequence of this is that for a very small size of the new phase, where a proportionally greater number of molecules are at the surface, the nucleus is unstable. At this stage, adding even a single molecule should only increase the free energy of the system, statistically leading the nucleus to dissolve. However, once the nucleus becomes large enough, the contribution of the bulk molecules overcomes the destabilisation in free energy induced by the surface molecules. The point at which the free energy of the system is decreased upon either growth or dissolution of the nucleus is called the critical size and is referred to as the Gibbs-Thomson effect.
Nucleation is the first real stage of crystallisation (after the necessary desolvation of molecules), with the occurrence of a phase transition and emergence of an organisation and an order in the system. A simple and comprehensive explanation of crystal formation has been given by Gibbs when he

mentioned two central driving forces: one promoting and the other one disfavoring the formation of a crystal. From a thermodynamical perspective, crystallisation can happen when experimental circumstances such as concentration, temperature or pressure cause the crystalline phase to be more stable than the corresponding melt or solution.[38] The work of Kashchiev *et al.*[37] suggests however that the most common scenario is that crystal formation can only occur if a certain free energy barrier is crossed.

In organic crystallisation, nucleation corresponds to the phenomenon of self-association of molecules in organised oligomers (monomers, dimers, trimers, *etc.*). The interactions between these emerging small systems and their environment (solvent molecules and solution additives) can significantly affect the formation of higher-order clusters during nucleation.

Two main theoretical frameworks have been developed to explain nucleation, the classical and the non-classical (or prenucleation cluster) nucleation mechanism (Figure 1.2).[39]

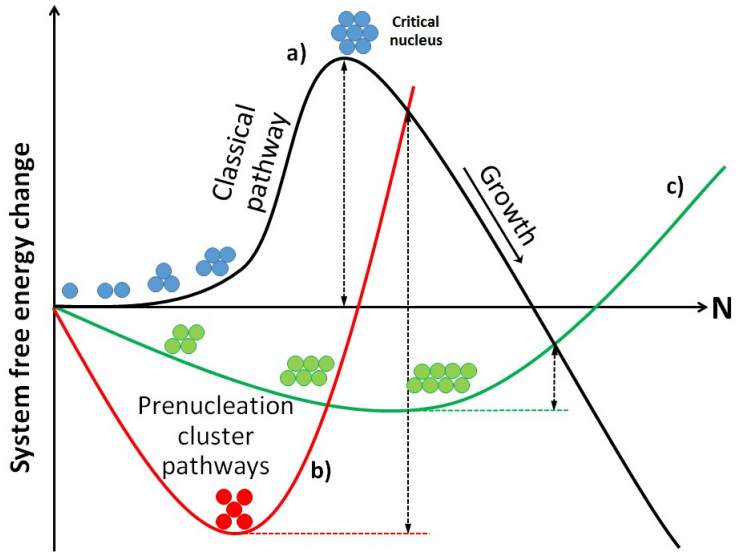

Figure 1.2: Free energy of the system as a function of crystal units. (a) Classical pathway. (b) and (c) Prenucleation pathways. PNCs of defined size increase the barrier to nucleation (b), whilst metastable disperse PNCs lower it, acting as building blocks to crystal growth (c).

According to *classical nucleation theory* (CNT), during the formation of clusters, the molecular packing reflects all possible polymorphs[40] and the crystal nuclei have the same structure as the mature crystal. For CNT, the aforementioned free energy barrier is linked to the need for a young crystal nucleus to form an interface with the solution. According to non-classical pathway, thermodynamically stable liquid-like clusters exist in solution and crystalline order originates from a separation of a disordered liquid phase.[41] A comparison of these two methods is displayed *figure 1.3*.

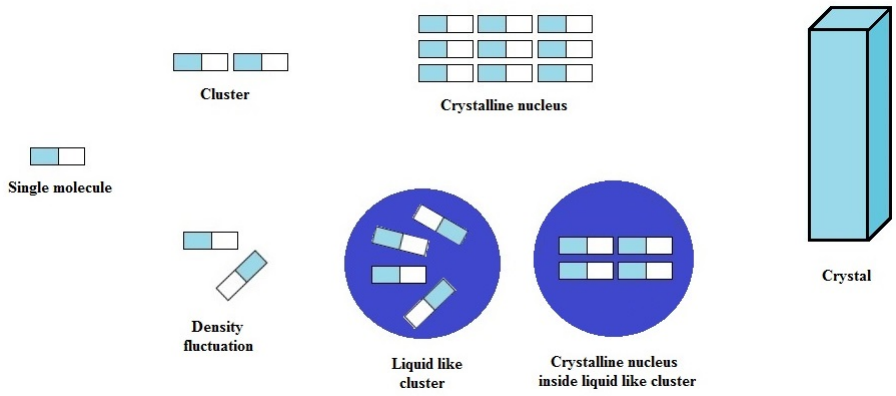

Figure 1.3: Nucleation pathways, adapted from Davey *et al.*[39]

Classical nucleation theory

The theoretical basis of classical nucleation theory is Gibbs' thermodynamics,[42] which considers the original liquid and the final crystal structure as the two only states involved in the phase transition. Crystallisation then occurs when experimental conditions (concentration, temperature or pressure) cause the solid-state phase to be more stable than the corresponding melt or solution.[38] CNT was developed by Becker, Döring and Volmer in the 1930s[42] and assumes that the nucleation process is a series of consecutive diatomic (or bimolecular) reactions:

$$A_1 + A_1 \rightarrow A_2$$
$$A_2 + A_1 \rightarrow A_3$$
$$\ldots$$
$$A_n + A_1 \rightarrow A_{n+1}$$

where A_1 is the single building unit (atom or molecule) and A_n corresponds to clusters consisting of n units. According to CNT, the free energy difference (ΔG) between a solution containing N dispersed building units

(A_1) and a system in which the N building units form a solid aggregate embedded by equal amounts of solvent molecules, which are at the same temperature and pressure conditions, is given by the sum of bulk and surface terms:

$$\Delta G = \Delta G_{bulk} + \Delta G_{surface} \tag{1.1}$$

For a spherical cluster A_n of radius R, these Gibbs free energy terms have the following form:

$$\Delta G = \frac{4}{3}\pi R^3 \rho_s \Delta\mu + 4\pi R^2 \gamma \tag{1.2}$$

where ρ_s is the number density of the bulk solid, $\Delta\mu$ is the difference in chemical potential between the liquid and solid, and γ is the interfacial free energy. If the solid phase is more stable than the liquid, the bulk term is negative, proportional to the volume of the cluster, and promotes the crystallisation. Because the surface molecules of the emerging phase are less bound compared to those in the solid-state phase, the surface free term is positive and disfavours the formation of the new solid-state phase. On the other hand, the bulk term is negative, proportional to the surface area of the cluster and favours the formation of the crystal.

Consequently, for small sizes of the new phase, where a greater number of molecules are at the surface, the nucleus is unstable. At this stage, adding even a single molecule should only increase the free energy of the system, statistically leading the nucleus to dissolve.[34] However, once the nucleus becomes sufficiently large, the contribution from the bulk molecules overcomes the destabilisation in free energy induced by the surface molecules

(Figure 1.2). The point at which the free energy of the system decreases upon either growth or dissolution of the nucleus corresponds to the critical radius:

$$r^* = \frac{2\sigma}{\rho_s \Delta\mu} \tag{1.3}$$

This is also known as the Thomson-Gibbs equation and gives the condition for equilibrium of the critical nucleus with the ambient phase. Beyond this threshold, aggregated nuclei are stable enough and free energy is released during the subsequent growth of the system (Figure 1.2).[43]

Prenucleation cluster mechanism

When investigating calcium carbonate, Gebauer *et al.* observed and reported evidence for stable $CaCO_3$ clusters, about 2 nm in diameter, that would instead grow by colliding and coalescing.[44] If this type of precritical clusters are to exist, their energy have to lie on a local minimum of the potential energy surface, which would only be possible if the classical assumptions are wrong.

In organic systems, evidences of PNCs have also been reported for amino acids, especially glycine, whose dimeric structure have been profusely debated over the past hundred years.[45–48] Early studies even suggested that the association of amino acids in solution may proceed beyond simple dimerisation.[49,50] Later on, utilising electrospray ionisation mass spectrometry,[51] large oligomers were observed for essentially any amino acid, independently of particular solution conditions,[52–55] confirming that the PNC crystallisation pathway is possible for organic molecules as well.

1.2.3 Growth

Growth is the process by which an already formed crystal becomes larger with the addition of more molecules or ions within its lattice. A first theoretical framework to explain crystal growth was proposed by Volmer with the development of the "layer-by-layer" growth model.[42]

Growth naturally follows nucleation and does not always differ in its chemical behaviour. The theory enunciated by Volmer assumed that growth of a specific crystal face was induced by the surface expansion following the formation of surface-bound two-dimensional nuclei that reached their critical size.

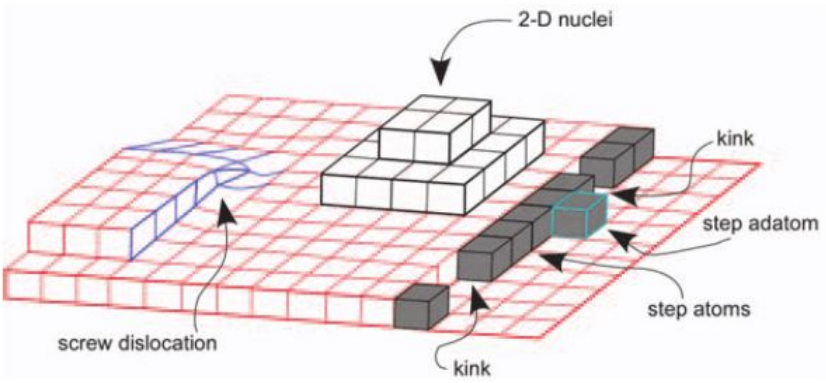

Figure 1.4: A schematic representation of classical growth theory from Teng[42]

Crystal growth from a supersaturated solution occurs because the flux of molecules that bind to the surface exceeds the flux of molecules leaving it. Consequently, special attention has to be drawn to the molecular processes at the solid-liquid interface, including the desolvation of the surface sites

and the adsorption of the solute molecules from the solution to the surface. In such processes, the specific interaction between the solvent molecules with the organic molecules at the surface and in solution plays a pivotal role in controlling the kinetics of crystal growth.[56] From a kinetic point of view, it is very interesting to note that, at conditions of equal supersaturations, highly soluble crystals generally will grow faster than sparingly soluble one; this has been explained in terms of the differences in the ratio of flux of molecules leaving the surfaces and flux of molecules attaching to the surface.[34]

An interesting example to cite is that of acentric crystals and the associated asymmetric growth.[57] Though the mechanism is not well understood, it was initially assigned mainly to solvent effect as well. However, later studies by Srinivasan and Sherwood[58,59] have shown that the unidirectional vapour growth cannot be primarily attributed to the solvent. The studies conducted by Singh *et al.* focused on growth and dissolution from vapour but are easily transferable and crystal grown from solution are known to further enhance the asymmetric character.[57]

A good example for the importance of crystallisation control in real life problems is that of calcium oxalate (CaOx) polymorphs, present as scale deposition on various equipments sensitive to energy losses.[60–62] In addition to polymorphism-related "issues", the very surface of a given compound is also an important parameter to consider, as it will behave differently for binding and interaction processes. The geometry of a surface is not the only factor that can influence subsequent growth, as it has been seen that molecules adsorbed on a surface could prevent a further growth of the covered surface.[63]

1.3 Solvent effect on polymorph control

Many mechanisms play a role towards the obtainment of a given polymorphic forms including growth in confinement, adjustment of the precipitating medium and formation of amorphous precursor phases.[64–70] However, it is the nature of the solution (type of solvent and presence of additives) that mostly influences polymorph selection during crystallisation from solution.[71] An example of the effect of solvent on the crystallisation process is provided by the *link-hypothesis*. Solution molecules tend to self-assemble in growth units and these aggregates often resembles the polymorph that crystallises from that particular solution, suggesting a *link* between the pre-nucleation aggregates and the resulting solid state structure.[72–75] Tetrolic Acid (TTA) is a good example of the importance of solvent effects, as the crystallisation from different environments (aqueous, alcoholic and apolar organic solvents) leads to the formation of different growth units, which structures have been observed in the final polymorph. *Figure 1.5* shows a qualitative explanation of the process of polymorph selection of two solvent-dependent crystal forms of tetrolic acid, α-TTA and β-TTA. In the *top* pathway, we have the formation of the so-called classic carboxylic dimer, which is stable in chloroform and corresponds to the structural synthon found in α-TTA. On the other hand, in ethanol (bottom panel), computer simulations have shown that the carboxylic dimer is thermodynamically unstable, TTA is present as single molecular units and the form that crystallises from solution is β-TTA.

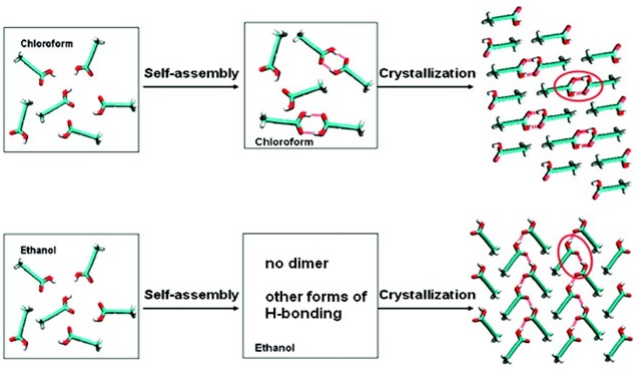

Figure 1.5: Tetrolic acid molecules in different solvents (left), in pre-nucleation aggregates (middle) and in their crystal form (right) from Chen *et al.*[76]

Several other systems show that the structural arrangement of clusters of solute molecules in solution affects the subsequent nucleation, including glycine[77] and 2,6-dihydroxylbenzoic acid[78]. However, there are also other systems such as mandelic acid[79] or anthranilic acid[80] where this connection between the stable molecular aggregates in solution and the crystal is absent. *Meta*-hydroxybenzoic acid also displays solvent-dependent crystallisation processes with the crystallisation of form I in ethanol and form II in ethyl acetate, but the classic carboxylic dimer motif is not present in the dominant polymorph that crystallises from the solvents in which the synthon is stable (ethyl acetate and acetonitrile).[81] Therefore, for carboxylic acids, the structure of the most stable species is not always linked to the pattern present in the polymorph that crystallises from solution. This goes against the link hypothesis and underlines the fact that the formation of a given form does not solely depend on the statistical occurrence of a specific shape of clusters.[82]

It has also been hypothesised that, when going from one polymorph

to another, and unless the solution is superheated for a long period of time, there might be some *memory effect* (as shown in Figure 1.6) and the compound could retain some ordering that influences nucleation (or renucleation) of a given polymorph.[83]

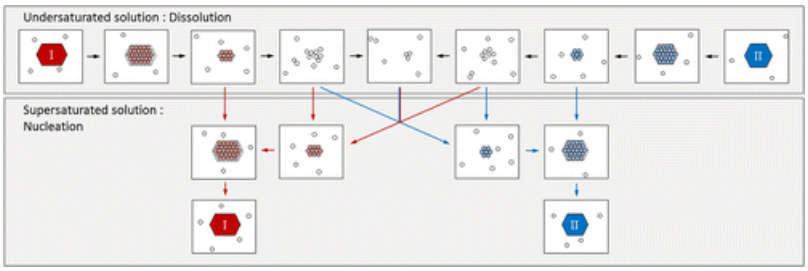

Figure 1.6: Renucleation pathways of molecules in solution starting from different polymorphs from Nordström *et al.*[83]

1.4 Previous computational studies

Significant advances have been achieved in the last decade in the field of crystal structure prediction to determine which polymorphs of an organic molecules can be found in its crystal energy landscapes.[84] It is worth mentioning the work conducted by Sally Price at UCL in the development of the crystal lattice simulation program DMACRYS, which is capable of distinguishing polymorphic energy differences to the accuracy of a few $kJ.mol^{-1}$.[85]

However, these methods exclude the role of solvent and do not consider the earliest stages of solute desolvation and crystal nucleation that, as previously discussed, can have a significant role in promoting the formation of a specific polymorph. Earlier work in this field was pioneered by Angelo Gavezzotti, who demonstrated by means of molecular dynamics (MD) simulations the aggregation of acetic acid in a carbon tetrachloride solution.[86]

However, due to computational constraints this study was limited to few nanoseconds of MD simulations. Using a similar methodology, Hamad and co-worked investigated the effect of concentration on the aggregation of glycine molecules in aqueous solution[87] and provided evidence for the formation of H-bonded clusters of found in the crystal structure of glycine. MD simulations give important information about the structure and population statistics of molecular clusters forming in solution, but they cannot explicitly provide quantitative data regarding the free energy change associated with the process of aggregation process.

In MD simulations, free energies can be computed by means of enhanced sampling techniques such as umbrella sampling or metadynamics. Both these methods force the exploration of regions of phase space along a defined collective variable that would otherwise have insufficient sampling in an unbiased MD simulation. In the umbrella sampling method, a harmonic bias potential $w_i = k_(x - x_i)^2$ along the collective variable x is used to explore energy regions around the target value x_i, and several simulations, differing only for their target values, are then run such that the statistics of neighbouring umbrellas overlap.[88] In metadynamics, configurations are automatically biased away from the most visited, and to escape local minima, the method adds periodically a Gaussian "hill" to the potential energy and slowly fills the potential well, forcing the system to explore high energy configurations.[89] Chen and Trout used MD umbrella sampling to compute the free energy profiles for the dimerisation of tetrolic acid in various solvents and rationalise the role of solvent in promoting the crystallisation of a dimer-based crystal.[76] More recently, remarkable work has been conducted by Salvalaglio and co-workers on the application of metadynamics methods

to determine the mechanism and of the role of the additives on the growth of urea surfaces.[90] However, except for simple molecular systems such as tetrolic acid and urea, the choice of collective variables used to drive the molecular self-assembly process in an enhanced sampling molecular dynamics simulation is not always obvious.
In this doctoral project, we have used both unbiased MD and metadynamics simulations to investigate the processes surrounding the crystallisation of meta-aminobenzoic acid, but we have also developed an alternative methodology to quantify the thermodynamic stability in solution of molecular clusters.

As both polymorph selection and crystallisation have been studied extensively in the past decades, it is always noteworthy to account for previous work in this field, especially from a computational point of view.
As one of the major barriers to the development of reliable computational models to predict the crystal structure is the uncertain number of polymorph, previous studies by Price *et al.* already highlighted the importance of computational methods while searching for polymorphs focusing both on kinetic and thermodynamic factors.[91] Notable simulations on both nucleation and growth of urea have been performed by Matteo Salvalaglio *et al.*, demonstrating the impact of the nature of the solution (solvent and additives) on the growth of selected surfaces[90] or highlighting the existence of a two-step mechanism (opposed to the more common CNT hypothesis) during nucleation.[92]
Similarly, remarkable work on nucleation theories has been conducted by Davey *et al.* A solid explanation on thermodynamics and kinetics of both

classical and non-classical nucleations theories were proposed,[39] along with a thorough case study on the phase transformation of a polymorphic system, 2,6-dyhydroxybenzoic acid.[39,93]

1.5 Meta-aminobenzoic

Meta-aminobenzoic acid (mABA) is of considerable importance in the pharmaceutical industry, widely used in the synthesis of analgesics, antihypertensives, vasodilators and other drugs.[94] This molecules also represents a fascinating model system for polymorphic research because it can crystallise in five different crystal structures (I-V).[95] The very strong polymorphic character of mABA can be related to its ability to form complex mABA/mABA inter-molecular interactions, such as hydrogen (H) bonding, π-π and H-π interactions, all of which are important generally during polymorph selection, but also to the fact that meta-aminobenzoic acid can be found in either or both the nonionic and zwitterionic forms (see figure 1.7).

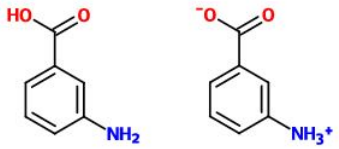

Figure 1.7: mABA molecule in its nonionic (left) and zwitterionic (right) forms.

In fact, mABA molecules are zwitterionic in the polymorphs denoted Forms I, III and IV, and nonionic in the other two polymorphs (Forms II and V)[96] (Figure 1.8). According to the recent work by Williams *et al.*, form III is more stable at ambient temperature than the previously found

forms I and II, showing that the polymorph of a crystal that is first discovered is not always the most stable one.[95] In solution the tautomeric form of mABA is solvent dependent, with mABA reported by different authors to be 50 % zwitterionic in water,[97,98] and as the nonionic form in acetonitrile, dimethyl sulfoxide (DMSO) and ethylacetate. Consequently, the crystal nucleation of the mABA polymorphs chiefly depends on the distribution of nonionic and zwitterionic molecules in solution. Forms I and III crystallise from water and methanol, whereas Form II is obtained in acetonitrile, dimethyl sulfoxide (DMSO) and ethylacetate[99]. Similarly to tetrolic acid and meta-hydroxybenzoic acid, mABA displays a solvent-dependent behaviour during crystallisation with the obtainment of form I in water and methanol and form II from acetonitrile.[99] It is also important to note that these two forms are enantiotropically related, that is they can change reversibly each into the other, by a temperature that was thermodynamically determined to be 156.1°C.[99]

It is worth mentioning that the two positional isomers *para*-aminobenzoic acid (pABA), a molecule that is involved in the production of benzocain, and *ortho*-aminobenzoic acid (oABA) also display polymorphism but to a much lesser extent than mABA. The pABA molecule can crystallise in two different forms, (namely α and β),[100] whereas for oABA three polymorphs have been found so far (labeled I, II and III).[101] With the only exception of form I of oABA[102], which contains a 1:1 ratio of non-zwitterionic/zwitterionic molecules, the constituent molecules of the other polymorphs of oABA and pABA are non-zwitterionic.

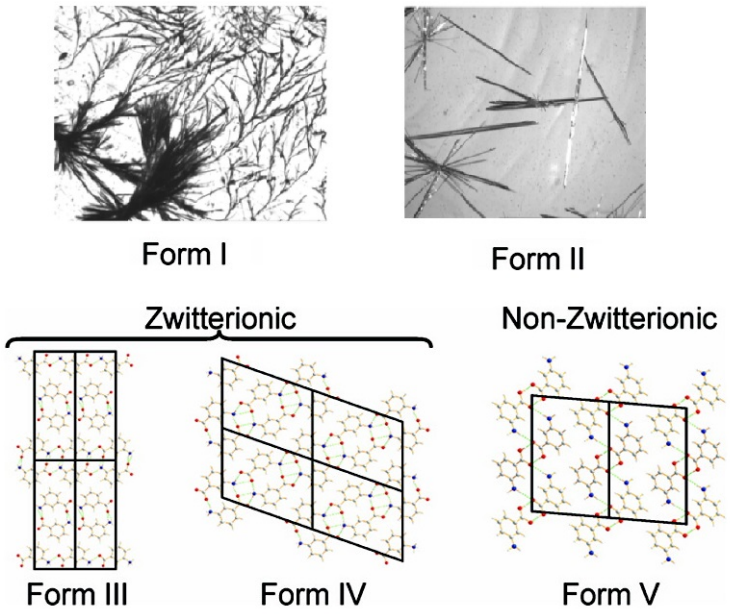

Figure 1.8: The five known polymorphs of mABA with zwitterionic (left) and non-zwitterionic (right) forms adapted from Svard *et al.*[99] and Williams *et al.*[95]

1.6 Aim of the project

Polymorph control during the formation of organic crystals from solution is determined by the solute-solvent interactions, making solution environments highly influential on the molecular processes controlling magnesite crystallization: (1) solute desolvation, (2) crystal nucleation, (3) surface growth.

The aim of this doctoral research project was the development and application of simulation techniques to resolve the atomistic phenomena driving the formation of meta-aminobenzoic acid (mABA), an active pharmaceutic ingredient important in polymorph research, including the role of solution chemistry in controlling the formation and growth of a specific polymorph of mABA.

The specific objectives that have been pursued during this project are associated with the three different steps involved in the crystallisation process:

- Solute desolvation: Quantify the structure and dynamics of the solvation structure around mABA molecules and the strength of interactions.
- Self-assembly of solute molecules: Characterise the thermodynamcs and kinetics of mABA self-association as a function of the solution environment.
- Surface reactivity and growth: Identity the molecular processes at solid-liquid interfaces driving mABA growth.

1.7 Thesis outline

This thesis reports the development and application of atomistic simulations to investigate the early stages of crystal nucleation and growth of mABA. Special attention has been given to the role of solution environment in controlling polymorph selection.

Chapter 2 provides a description of the theoretical methods used in the present thesis, which comprised a combination of electronic structure calculations and interatomic potential forcefield methods. Details of the computational chemistry codes used, simulation protocols adopted and algorithms developed during the course of the project are also presented.

Chapter 3 discusses the behaviour of the zwitterionic and non-zwitterionic mABA molecule in DSMO or water to identify the preferred ionic state of

mABA in solution, study the solvation structure of mABA and quantify the solvent-solute strength of interaction.

Chapter 4 focuses on the aggregation of small low-lying structure of non-zwitterionic mABA oligomers. By means of quantum chemistry and molecular dynamics, the quantification of the thermodynamics and energetics of these $(\mathrm{mABA})_{\mathrm{n}}$ systems (n=2-4) is provided both in the gas and solution-phase.

Chapter 5 considers the zwitterionic form of mABA, which is present in three of the five polymorphic forms, and on complex mixtures of mABA/$\mathrm{mABA}^{\pm}$.

Chapter 6 presents calculations on the stability of low-indices surfaces obtained from non-zwitterionic form II of mABA. Extensive molecular dynamics simulations on the selected stable surfaces in contact with a solvated environment are then analysed to understand the behaviour at the interface between organised mABA systems and the solvent.

Finally, chapter 7 provides a summary of all the results obtained during this project and scope for future work.

Chapter 2

Theoretical methods

This Chapter provides a description of the theoretical methods used in the present thesis. The first two sections discuss the methods adopted to compute the energy of the system using electronic structure calculations (density functional theory) and interatomic potential forcefield methods. In the third section the molecular dynamics technique is extensively described. The chapter then describes the methods used to treat the solvation environment using implicit continuum polarisable methods.
Finally, we give an overview of the computational chemistry codes, simulation protocols and algorithms that have been used and developed during this doctoral project.

2.1 Density functional theory

Density functional theory (DFT) is a quantum mechanical method to solve the time-independent, non-relativistic Schrödinger's equation:

$$\hat{H}\Psi_i = E\Psi_i \tag{2.1}$$

where Ψ_i is the wave function and $\hat{\mathrm{H}}$ is the Hamiltonian of a system comprising M nuclei and N electrons. The Hamiltonian is defined as sum of the following energy operators:

$$\begin{aligned}\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \\ + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}\end{aligned} \tag{2.2}$$

where the capital letters refer to the M nuclei and the lower case letters refer to the N electrons. In equation 2.2, $\mathrm{Z_A}$ and $\mathrm{Z_B}$ are the atomic numbers of the nuclei A and B, $\mathrm{r_{ij}}$ is the distance between electrons i and j and $\mathrm{R_{AB}}$ is the distance between nuclei A and B. The equation is divided in five distinctive energetic components. From left to right, we have the electronic kinetic contribution, the nuclear kinetic contribution, the electron-nucleus attractive contribution, the electron-electron repulsive contribution and the nucleus-nucleus repulsion contribution. The first approximation to be considered was described by Born and Oppenheimer (BO) and states that, due to their masses, the nuclei are much slower than the electrons and can therefore be considered as fixed compared to the movements of the electrons. Because of the BO approximation, nuclear kinetic energy is zero and the potential energy becomes a constant. The electronic Hamiltonian can therefore be written as:

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \qquad (2.3)$$

The solution of the Schrödinger equation with the electronic Hamiltonian is the electronic wave function and the electronic energy according to the equation:

$$\hat{H}_{elec}\psi_{elec} = E_{elec}\psi_{elec} \qquad (2.4)$$

The electronic wave function of a N-electrons molecule depends on 3N spatial and N spin coordinates, and cannot be solved exactly. An efficient and accurate approach to compute, indirectly, the electronic energy of a set of interacting electrons is DFT, which is based on the work of Hohenberg and Kohn, who proved that the ground state electronic energy is completely defined by the knowledge of the ground-state electron probability density:

$$\rho(\vec{r}) = N\int ...\int |\psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_N)|^2 ds_1 d\vec{x}_1 ... d\vec{x}_N \qquad (2.5)$$

where $\rho(\vec{r})$ is the probability of finding any of the N electrons within the infinitesimal volume $d\vec{r}$. The electron density is a function of only the three spatial variables, and satisfies the following conditions:

$$\begin{cases} \rho(\vec{r} \to \infty) = 0 \\ \int \rho(\vec{r}) d\vec{r} = N \end{cases} \qquad (2.6)$$

Equation 2.6 highlights the need to conserve the number of particles. Integration of the electron densities on the entire space domain renders the

total number of particles and, at an infinite distance, the electron density is zero.

2.1.1 Kohn-Sham approach

The most common implementation of DFT is based on the Kohn-Sham (KS) scheme. This approach considers a fictitious system of non-interacting electrons having, however, the same electronic density and consequently energy as the real system. The electronic Hamiltonian of a system consisting of N non-interacting electrons can be written as the sum of single particle operators:

$$\hat{H}_e^{KS} = \sum_{i=1}^{N}[-\frac{1}{2}\nabla_i^2 + V_{eff}(\vec{r_i})] \equiv \sum_{i=1}^{N}\hat{h}_{eff}^{KS}(i) \tag{2.7}$$

where $V_{\text{eff}}(\vec{r})$ is the effective Kohn-Sham potential. The Schrödinger equation for this fictitious system is

$$H_e^{KS}\Psi_0^{KS} = E_0^{KS}\Psi_0^{KS} \tag{2.8}$$

and its solution can be written exactly as a single Slater determinant:

$$\Psi_0^{KS} = \frac{1}{\sqrt{N_e}}|\psi_1, \psi_2, ..., \psi_{N_e}| \tag{2.9}$$

where

$$\hat{h}_{eff}^{KS}\psi_i = \epsilon_i\psi_i \tag{2.10}$$

is the single particle Schrödinger equation that determines the single-particle orbitals ψ_{i}, under the constraint that the orbitals ψ_{i} are orthonor-

mal, $\langle\ \psi_i\ |\ \psi_j\ \rangle = \delta_{\mathrm{ij}}$. The electron density of this fictitious system of non-interacting electrons is given by:

$$\rho(\vec{r}) = \sum_{i=1}^{N_e} |\psi_i(\vec{r})| \tag{2.11}$$

where the summation runs over the N occupied orbitals. It is possible to show that the effective Kohn-Sham potential in Eq. 2.7 has the following general expression:

$$V_{eff}(\vec{r}) = V_{ext}(\vec{r}) + \int \frac{\rho(\vec{r'})d\vec{r'}}{|\vec{r} - \vec{r'}|} + \nu_{xc}[\rho(\vec{r})] \tag{2.12}$$

where $V_{ext}(\vec{r}) = \Sigma_{\mathrm{Ij}} \frac{Z_I}{|\vec{R_I} - \vec{r}|}$ is the external potential acting on the electron due to the nuclei as the positions $\{\vec{R_I}\}$; $V_H(\vec{r}) = \int \frac{\rho(\vec{r'})d\vec{r'}}{|\vec{r} - \vec{r'}|}$ is the Hartree potential; $\nu_{\mathrm{xc}}[\rho(\vec{r})]$ is the local exchange-correlation potential. The single particle Kohn-Sham equations are therefore:

$$\left[-\frac{1}{2}\nabla^2_{\vec{r_i}} + V_{ext}(\vec{r}) + V_H(\vec{r}) + V_{xc}(\vec{r})\right]\psi_i = \epsilon_i\psi_i \tag{2.13}$$

and it can be shown that the ground-state electronic energy of the system is given by

$$E_O^{KS} = -\frac{1}{2}\sum_{i=1}^{N_e}\langle\psi_i|\nabla^2_{\vec{r_i}}|\psi_i\rangle + \int V_{ext}(\vec{r})\rho(\vec{r})d\vec{r} + \frac{1}{2}\int V_H(\vec{r})\rho(\vec{r})d\vec{r} + E_{xc}[\rho(\vec{r})] \tag{2.14}$$

According to Eq. 2.14, $\mathrm{E_0}^{\mathrm{KS}}$ is a functional of the electronic density $\rho(\vec{r}) = \Sigma_{\mathrm{i=1}}{}^{\mathrm{N_e}}\ |\ \psi_{\mathrm{i}}(\vec{r})\ |$

2.1.2 Jacob's ladder

The exact form of the exchange-correlation is not known and several levels of approximations have been developed for the $E_{xc}[\rho(r)]$ term in Eq. 2.14. A schematic ranking of all families of approximations has been proposed by Perdew and is known as the *Jacob's ladder* in Fig. 2.1.

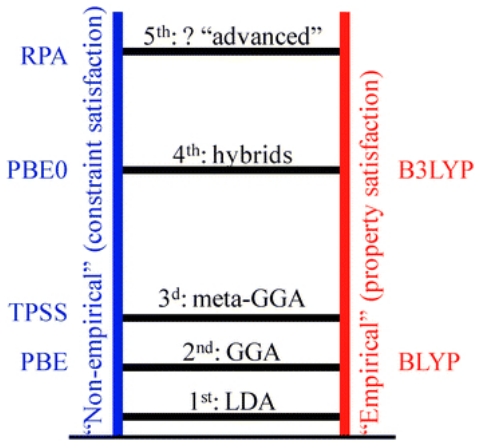

Figure 2.1: Representation of *Jacob's ladder* density functional approximations for the exchange correlation term by Perdew *et al.*[103]

The most common types of approximations are Local Density Approximation (LDA) and Generalized Gradient Approximation (GGA). LDA approaches a real system by dividing it into infinitesimal volumes where the electron density is considered to be a constant:

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r})\epsilon_{xc}^{unif}(\rho(\vec{r}))d\vec{r} \tag{2.15}$$

where ϵ_{xc}^{unif} is the exchange-correlation energy per particle and $\rho(\vec{r})$ is the electronic density.

However, the local uniform density at each point is not a reasonable approximation for the rapidly varying electronic density of certain materials.

It was later found that the inclusion of the gradient of the density yields better results for many properties of molecular and solid state systems:

$$E_{xc}^{GGA}[\rho] = \int f^{GGA}(\rho(\vec{r}), \nabla\rho(\vec{r}))d\vec{r} \tag{2.16}$$

One of the main limitations of standard exchange-correlation density functionals is their inability to fully describe dispersion interactions, which can affect the accuracy when computing the structural and energetic properties of liquids, molecular complexes and organic solid state materials. In the last decade, several new, modern functionals have therefore been developed to overcome this shortcoming by LDA, GGA and hybrid functionals, that include a combination of exact exchange from Hartree-Fock theory enhanced with exchange-correlation energy from other sources. For example, the B97-D method has been developed by Grimme *et al.*[104] and is based on the B97 gradient-corrected functional to which has been added an empirical dispersion correction ($\frac{C}{r^6}$), where C is the dispersion parameter.

2.1.3 Finding the minimum

According to the Raileigh-Ritz variational principle, the ground state electronic wavefunction Ψ_0 is the wavefunction that minimises the expectation value of the electronic Hamiltonian:

$$E_o(\{\vec{R_I}\}) = \min_{\Psi_0}\langle\Psi_0|H_e(\vec{R_I})|\Psi_0\rangle \tag{2.17}$$

In the framework of the Kohn-Sham DFT approach, $\Psi_0 = \det\{\psi_i\}$ (Eq. 2.9) and the ground state electronic energy of the fictitious system can be obtained by minimising the expectation value $\langle\Psi_0^{KS}|H_e^{KS}(\vec{R_I})|\Psi_0^{KS}\rangle$ with

respect to the orbitals ψ_i:

$$E_0^{KS}(\{\vec{R}_I\}) = \underset{\{\psi_i\}}{min}\langle\Psi_0|H_e(\vec{R}_I)|\Psi_0\rangle \tag{2.18}$$

subject to the constraint that the orbitals ψ_i are orthonormal, $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. In the standard implementation of DFT, the Kohn-Sham single-particle orbitals$\{\psi_i\}$ are expanded in terms of a basis set $\{\phi_\alpha(r), \alpha = 1, N_\alpha\}$:

$$\psi_i(\vec{r}) = \sum_\alpha c_\alpha^{(i)}\phi_\alpha(\vec{r}) \tag{2.19}$$

where $\{c_\alpha^{(i)}\}$ are the expansion coefficients for the i-ith orbital. In the solid state community, the most popular choice is the *plane wave* basis set, whereas many traditional quantum chemistry programs such as Gaussian, Turbomole and ADF use atom-centered (Gaussian or Slater) functions. The expansion (2.19) for the orbitals can be substituted in the Equation (2.14) to provide an expression of the electronic energy as a function of the expansion coefficients $\{c_\alpha^i\}$:

$$E_0^{KS} = E_0^{KS}\Big(\{c_\alpha^{(i)}\}\Big) \tag{2.20}$$

For a fixed set of nuclear positions $\{\vec{R}_I\}$ (direct minimisation of the energy) the minimum condition (2.17) can be seen as an optimisation problem for the cost function $E^{KS}\Big(\{c_\alpha^{(i)}\}\Big)$ with respect to the multitude of coefficients $\{c_\alpha^{(i)}\}$

$$E_0^{KS}(\{\vec{R}_i\}) = \underset{\{c_\alpha^{(i)}\}}{min}E^{KS}\Big(\{c_\alpha^{(i)}\};\{\vec{R}_I\}\Big) \tag{2.21}$$

A simple strategy to solve this optimisation problem is to: (i) com-

pute the gradient $\delta E^{KS}/\delta c_\alpha^{(i)}$ of the energy with respect to the expansion coefficients; (ii) exploit this information in a steepest descent or conjugate gradient routine to approach the minimum. The practical feasibility of this approach depends on the following two requirements:

1. The evaluation of the gradient should be inexpensive, which is the case, for example, when the basis set used are plane waves, where $\delta E^{KS}/\delta c_\alpha^{(i)}$ can be computed efficiently exploiting the fast Fourier transform.

2. The hypersurface $E^{KS}\Big(\{c_\alpha^{(i)}\}\Big)$ has to be simple, *i.e.* it should not present local minima preventing the system from reaching the ground state. The experience has shown that the surface $E^{KS}\Big(\{c_\alpha^{(i)}\}\Big)$ is indeed very simple, but a critical step is the availability of a good starting approximation for the required eigenvectors, *i.e.* a good starting set of expansion coefficients $\{c_\alpha^{(i)}\}$.

2.2 Forcefield

A forcefield is a mathematical expression, also known as potential function, for the potential energy of a molecular system that depends parametrically on a set of constants.

For organic molecules, one widely used family of interatomic potential models is the Assisted Model Building with Energy Refinement (AMBER) force field. AMBER describes the energy of a molecular system as the sum of five decoupled energy functions that are presented in equation 2.22 and figure 2.2:

$$
\begin{aligned}
E_{\text{total}} = & \sum_{\text{bonds}} K_r(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\
& + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{paired} \frac{q_i * q_j}{4\pi\epsilon_0 r_{ij}} \\
& + \sum_{unpaired} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]
\end{aligned}
\tag{2.22}
$$

From left to right and top to bottom, the five terms of this equation are:

- Bonds' energy: an Harmonic functional form where K_r and r_0 are the harmonic force constant and reference bond length parameters, respectively, and r is the distance.

- Angles' energy: K_θ is the harmonic constant for the angular three-body term, θ is the angle and θ_0 is the reference angle.

- Dihedrals' energy: V_n is the torsional constant, ϕ is the current torsional angle and γ is the phase angle.

- Paired atoms' energy: q_i and q_j are the partial charges assigned to atoms i and j, ϵ_0 is the vacuum permittivity and r_{ij} is the distance between particles i and j. This term is the expression of the Coulomb's law for separated electric charges.

- Unpaired atoms' energy: ϵ_{ij} is the depth of the potential well, σ_{ij} is the finite distance at which the potential is zero and r_{ij} is the distance between particles i and j. This term is the expression of the Lennard-Jones potential energy.

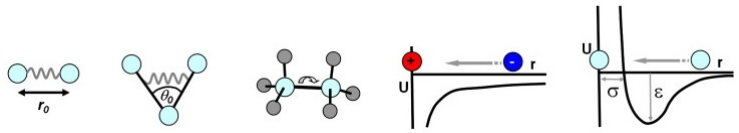

Figure 2.2: From left to right, visualisation of the force-field parameters for bonds, angles, torsions, electrostatic and Van der Waals energies.

A general set of parameters for small organic molecules compatible with the AMBER forcefield is provided by the GAFF database, which has parameters for almost all organic molecules containing C, N, O, H, S, P, F, Cl, Br and I. The GAFF forcefield was also previously developed to simulate the solvation, aggregation and crystal growth of several organic molecules, including p-aminobenzoic acid.[105]

2.3 Molecular dynamics

Molecular dynamics (MD) is a simulation technique to solve the equations of motion for a set of interacting particles. MD allows to follow the time evolution of complex molecular systems such as solutions and solid-liquid interfaces, and from their trajectory to compute the structural, dynamic and thermodynamics properties of the system.

2.3.1 The equation of motion

In Newtonian mechanics, the trajectories (position and momentum as a function of time) of the atoms of a molecular system are determined by solving the Newton's equations of motion:

$$\begin{cases} \frac{d\vec{r_i}(t)}{dt} = \frac{\vec{p_i}(t)}{m} \\ \frac{d\vec{p_i}(t)}{dt} = \vec{F_i} \end{cases} \tag{2.23}$$

Where $\vec{p}(m)$=$m\vec{v}(t)$ is the momentum of the particle with mass m and $\vec{F_i} = \vec{F_i}(\vec{r_1}, \vec{r_2}, ..., \vec{r_N})$ is the force acting on a given atom depending on the position of all other atoms.

The equations of motions cannot be solved analytically because of the functional form of the force acting on each particle also depends on the position of all other particles. The simulation technique of molecular dynamics solves these differential equations using a finite difference approach. The general idea is to break the calculation into a series of very short time steps (usually around the *femtosecond*): (i) at each step, the force on the atoms are computed and combined with the actual positions and velocities to generate the new positions and velocities one time step (Δt) ahead; (ii) the force acting on each atom is assumed to be constant during each interval; (iii) the particles are then moved to the new positions and a new set of forces is computed; (iv) the process is repeated. Using this algorithm, a MD simulation generates a dynamical trajectory describing the time evolution of the positions $\vec{r}$ and momenta $\vec{p}$ of each particle in the system.

MD is therefore a simulation technique to follow the microscopic dynamics of a system, which is the trajectory of each particle within it. MD is also a *deterministic* technique because, in principle, given the initial set of positions $\mathrm{r_i(t_0)}$ and momenta $\mathrm{p_i(t_0)}$, the subsequent time evolution is completely defined.

2.3.2 Statistical mechanics

Statistical mechanics is a branch of physical sciences providing the exact mathematical expressions linking macroscopic thermodynamic properties to the microscopic details of the system such as the position and velocities of the particles in a N-body system.

The *thermodynamic state* of a system is defined by a small set of parameters, such as temperature T or pressure P, while the *microscopic state* is defined by $\mathrm{x^N} = \{\mathrm{r^N},\mathrm{p^N}\}$, where r^N and p^N are the atomic positions and momenta of all N particles of the system. A statistical *ensemble* a collection of all possible systems which differ in their microscopic state but have the same thermodynamic state. In other words, we are looking at many, individual microscopic configurations of a very large system leading to the same macroscopic properties. There are different ensembles with different characteristics:

- The *microcanonical* ensemble **NVE** is defined by a constant number of moles **N**, a fixed volume **V** and energy **E**.
- The *canonical* ensemble NVT is characterised by a constant number of moles **N**, a fixed volume **V** and temperature **T**.
- The *isothermal-isobaric* ensemble NPT regroups systems which thermodynamic state is ruled by constant number of moles **N**, a fixed volume **V** and pressure **P**.
- The *grand-canonical* ensemble νVT, that we did not use in this present work, is a collection of systems for which the thermodynamic state is characterised by a fixed potential ν, a fixed volume V and a

fixed temperature P.

Because the energy is conserved during a simulation, MD generates configurations in the NVE ensemble but techniques are available to sample configurations in other ensembles.

2.3.3 Modelling the physical system

A critical part of a MD simulation is how to model the interaction between the particles of the molecular system under study. This is related to the choice of the potential energy function, that is a function of the positions of the nuclei (electrons are still not explicitly taken into account) $\mathrm{V}(\vec{r_1},\vec{r_2},...,\vec{r_\mathrm{n}})$, corresponding to the potential energy of the system when the atoms are in that specific configuration. In classical MD the potential energy function is given by a forcefield such as the AMBER one. Instead, in *ab initio* MD, the potential energy function corresponds to the expectation value of the electronic Hamiltonian and is usually obtained by solving the Kohn-Sham equations (DFT). The forces are then derived as the gradients of the potential with respect to the atomic displacements:

$$\vec{F_\mathrm{i}} = -\nabla_{\vec{r_i}} V(\vec{r_1}, \vec{r_2}, ..., \vec{r_n}) \tag{2.24}$$

2.3.4 Integration algorithms

The engine of a MD code is the algorithm required to integrate numerically the equation of motion of the interacting particles and compute their trajectory (positions and velocities at each time step). Integration algorithms are based on finite difference methods with a "discretised" time on a finite

grid on which the time step Δt is the distance between consecutive points (the exact details depend on the type of algorithm). By iterating the procedure, the time evolution of the system can be followed for a certain period of time.

The most popular integration methods for MD simulations are based on the Verlet algorithm (Verlet, leap-frog and velocity Verlet). The basic idea of the Verlet algorithm is to write two third-order Taylor expansion for the position vector $\vec{r}$(t), one forward and one backward in time (t+Δt and t-Δt). We obtain:

$$\begin{cases} \vec{r}(t+\Delta t) = \vec{r} + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 + \frac{1}{6}\vec{b}(t)\Delta t^3 + O(\Delta t^4) \\ \vec{r}(t-\Delta t) = \vec{r} - \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 - \frac{1}{6}\vec{b}(t)\Delta t^3 + O(\Delta t^4) \end{cases} \tag{2.25}$$

Adding these two equations gives:

$$\vec{r}(t+\Delta t) = 2\vec{r}(t) - \vec{r}(t-\Delta t) + \vec{a}(t)\Delta t^2 + O(\Delta t^4) \tag{2.26}$$

This is the basic form of the Verlet algorithm, with $\vec{v}$ the velocities (first time derivative of the position), $\vec{a}$ the accelerations (second time derivative of the position) and $\vec{b}$ the third derivative with respect to time. Due to their opposite signs, velocities have been eliminated in the addition process. Velocities are not necessary for generating trajectories but they are require to compute the kinetic energy (K) and the total energy of the system (E=K+V). If needed, velocities are obtained from the positions using:

$$\vec{v}(t) = \frac{\vec{r}(t+\Delta t) - \vec{r}(t-\Delta t)}{2\Delta t} \tag{2.27}$$

In this study, we used the leap-frog method that is quite similar to the velocity Verlet algorithm. In leap-frog integration, the position and velocity are given by:

$$\begin{cases} \vec{x}_i = \vec{x}_{i-1} + \vec{v}_{1-1/2}\Delta t \\ \vec{a}_i = F(\vec{x}_i) \\ \vec{v}_{i+1/2} = \vec{v}_{i-1/2} + \vec{a}_i\Delta t \end{cases} \tag{2.28}$$

where $\vec{x}_i$ is the position at step i, $\vec{v}_{i+1/2}$ is the velocity at step $i+1/2$, $\vec{a}_i$ is the acceleration at step i and Δt is the size of each time step. Positions and velocities can also be expressed at integer steps:

$$\begin{cases} \vec{x}_{i+1} = \vec{x}_i + \vec{v}_i\Delta t + \frac{1}{2}\vec{a}_i\Delta t^2 \\ \vec{v}_{i+1} = \vec{v}_i + \frac{1}{2}(\vec{a}_i + \vec{a}_{i+1})\Delta t \end{cases} \tag{2.29}$$

It is to note that the leap-frog integration is often used in Hamiltonian Monte-Carlo because of its *symplectic nature* (conservation of the energy of dynamical systems) and its *time-reversibility*.

Time step

The time step (Δt) used in the integration algorithm is a critical parameter when performing MD simulations. The choice of Δt represents a compromise between accuracy and cost of the simulation. There is not a simple and unique recipe to determine the most appropriate time step for a molecular dynamics simulation. If the time step is too small, the trajectory will only cover a limited portion of the phase space; if the time step is too large, it leads to instabilities in the integration algorithm due to high energy over-

laps between atoms. As often with computational techniques, the aim is to find a correct balance between simulating the trajectory and covering the phase. In the present work, the time step was set around 1 fs.

2.3.5 Periodic boundary conditions

An important technical aspect when simulating liquids or solids is represented by the boundaries of the system. If nothing is done, the system simply terminates and atoms near the boundary will have fewer neighbours than the atoms inside, just as if the sample would be surrounded by surfaces. It is computationally unfeasible to simulate a system with a number of atoms comparable to a macroscopic system, because this simplification would lead to surface effects being much more important than what they should. A solution to this problem is to employ *periodic boundary conditions* (PBC).

When using PBC, particles are enclosed in a box that is modelled to be replicated to infinity by rigid translation in all three cartesian directions. If a particle has a position given by $\vec{r}$, we assume it actually really represents a set of particles located at:

$$\vec{r} + l\vec{a} + m\vec{b} + n\vec{c}(l, m, n = -\infty, +\infty) \tag{2.30}$$

where l, m and n are integers and $\vec{a}$, $\vec{b}$ and $\vec{c}$ are the vectors corresponding to the edges of the simulation box. As a result, whenever an atom leaves the simulation cell, it is replaced by another with exactly the same velocity, entering from the opposite cell face. The number of atoms in the cell is conserved and no atom feels any surface forces, removing the boundary

problem.

2.4 Computational procedures

This section provides a brief description of the computational chemistry codes used during this project to simulate the solvation, aggregation and surface growth of mABA crystallisation from solution. The algorithms and software tools developed to generate molecular models and analyse results will also be considered.

2.4.1 Simulation software

- **Gaussian** (version 09) is a commercial molecular electronic structure code used to conduct ground state as well as excited state calculations in the framework of Hartree-Fock (HF), post-HF and DFT methods using as basis set Gaussian-Type Orbitals (GTO). The Gaussian code includes a large number of exchange-correlation functionals including DFT methods that accurately describe long-range van der Waals interactions, such as M06-2X and B97-D3. Moreover, several polarisable continuum solvation models have been implemented in the Gaussian code.

  In particular, universal Solvation Model based on Density (SMD) developed by Truhlar and co-workers[106] has been used in this project to evaluate the solvation free energies of monomers and oligomers of mABA in aqueous and organic solutions. SMD is recommended in Gaussian for the calculation of solvation free energies (ΔG) of solvation because its assessment of over a set of 2892 solvation free

energies and transfer free energies for neutral solutes and ions in water and non-aqueous solutions gave a mean unsigned error (MUE) of only 0.8 kcal.mol^{-1} for neutral solutes, and 4.3 kcal.mol^{-1} for ions. In particular, the MUE in calculated solvation free energies of carboxylic acids was only 0.25 kcal.mol^{-1} in water and 0.55 kcal.mol^{-1} in organic solvents using the M06-2X/6-31+G(d,p) level of theory. The SMD model together with the M06-2X density functional was also applied to predict the free energies of aqueous solvation for 61 drug-like molecules in the SAMPL1 test set and the authors reported a MUE of 2.0 kcal.mol^{-1}.[107] The SMD/M06-2X level of theory was also used by Ribeiro and co-workers to compute the partition coefficients of nucleobases between chloroform and water with a MUE of 0.8 kcal.mol^{-1}.[108]

- **NWChem** (version 4.3) is a molecular electronic structure code that, similarly to the Gaussian program, uses GTOs basis set functions. NWChem does not have as many solvation models of density functionals as Gaussian but it has been specifically designed to run on high performance computing (HPC) systems. Moreover, the structure of the input file allows for the calculation of sequential jobs.

  Therefore, NWChem has been adopted in this project to conduct high-throughput evaluation of the energies of candidate structure and to perform preliminary geometry optimisation. NWChem calculations have been conducted on the Queen Mary cluster as well as on the UK supercomputing facility (Archer).

- **Gromacs** (GROningen MAchine for Chemical Simulation version 5.0.4) is a molecular dynamics code that was initially mainly de-

veloped to conduct simulations of biological systems (proteins, lipids and macromolecules). However, in this project, Gromacs has been used to conduct simulations of the aggregation and surface reactivity of mABA in contact with water and DMSO.

In addition, a version of GROMACS equipped with the PLUMED plugin[109] has been used to conduct metadynamics simulations.

- **CP2K** (version 4.1) is a freely available (GPL) program to perform calculations of molecules, solids and liquids using both electronic structure and forcefield methods. The uniqueness of CP2K is the use of a mixed Gaussian and plane waves approach (GPW).

  CP2K has been mainly used to conduct *ab initio* (Born-Oppenheimer) molecular dynamics simulations. Test calculations have shown that AIMD conducted using CP2K is twice as fast as simulations conducted with traditional plane-wave codes such as VASP or Quantum-ESPRESSO.

- **Quantum Espresso** (quantum opEn-Source Package for Research in Electronic Structure, Simulation, and Optimization version 6.3) is a traditional software based on density functional theory, plane waves, and pseudopotentials.

  In this project, it has been used to conduct variable-cell optimisation of the unit cell of form II of mABA and benchmark calculations for CP2K.

- **Granada** is a code designed to generate molecular structures by distributing one or more mobile molecule around a central molecular unit (monomer, dimer, trimer, *etc.*).[110] Each *mobile* molecule is ran-

domly situated in a new centre of coordinates with respect to the *central* molecule and is also randomly oriented along the three coordinate axes. It was initially developed by Montero *et al.* to generate micro-hydrated molecular structures.[111]

In this project, it has been used to generate configurations for dimers, trimers and tetramers as well as configurations for monomers in contact with explicit solvent.

2.4.2 In-house codes

Scripts and codes written in Bash and FORTRAN (FORmula TRANslator) languages were developed to automatise the preparation of input files, submission of calculations and subsequent data gathering from the output files.

Configuration selector: The Granada code generates a defined number of random configurations within a box of given size. However, some of the candidate configurations are unlikely to be of significance because one or more of the randomly mobile molecules are too far away between each other or from the central unit (molecule or oligomer) placed at the center of the box. Using simple geometrical restrictions, the Configuration Selector program is a tool to select plausible configurations generated by Granada.

The input is the list of configurations generated by Granada (usually hundreds of thousands) in the XYZ format. The algorithm has a double loop that runs over all the atoms of the molecular cluster. If one atom of a mobile molecule is within the set threshold, the molecule is accepted. If all the molecules are accepted, the configuration is accepted. It has been

used for the generation of small clusters from building blocks (monomer) and explicitly solvated systems.

GRO2GROW: this short program is used to rearrange an output file generated by *cif2cell* into a Gromacs readable input file. As the .gro output from the *cif2cell* package is not exactly formatted for Gromacs, this intermediate software tool has been developed for the generation of .gro configuration input files.

COMFORT (Centre Of Mass in FORTran): this program reads an output file from a Gromacs molecular dynamics simulation and creates a file containing the coordinates for the centres of masses of each molecules at every step of the simulation. Two versions of the codes were developed: one taking into account Periodic Boundary Conditions (PBC) (*equations 2.32 to 2.35*) and one not (*equation 2.31*).
The program loops over the lines in the input file, reading the name and the coordinates for each atom and assigning a value for the molecular mass. Then it calculates the coordinates for the centre of mass according to the equation for non-PBC:

$$\vec{R} = \frac{1}{M} \sum_{i=1}^{n} m_i \vec{r}_i \tag{2.31}$$

where $\vec{R}$ is the position vector of the centre of mass, M is the total mass of all particles, m_i is the mass of the particle i and $\vec{r}_i$ is the position vector of particle i.

The set of equations to define centre of mass in PBC is:

$$\theta_i = \frac{x_i}{x_{max}} 2\pi \tag{2.32}$$

where x_i is the current position and x_{max} is the maximum size following one direction.

$$\begin{cases} \xi_i = cos(\theta_i) \times m_i \\ \zeta_i = sin(\theta_i) \times m_i \end{cases} \tag{2.33}$$

where m_i is the mass of particle i

$$\overline{\theta} = atan2(-\overline{\zeta}, -\overline{\xi}) + \pi \tag{2.34}$$

$$x_{com} = x_{max}\frac{\overline{\theta}}{2\pi} \tag{2.35}$$

where x_{com} is the x coordinate for the centre of mass.

Chapter 3

Solvation of meta-aminobenzoic acid in aqueous and organosulfur solutions

As the desolvation of the monomers of meta-aminobenzoic acid represents the first step in the crystallisation process, the intermolecular interactions between the solute and the surrounding solvent molecules can have a significant influence on the kinetics of organic crystal formation. Meta-aminobenzoic acid, an important model system in the study of polymorphism of active pharmaceutical ingredients, can exist in both the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) forms, where its distribution in solution depends chiefly on the nature of the solvent. In this chapter, atomistic simulations were used to study the structural and energetics of the monomers of meta-aminobenzoic acid in aqueous and organosulfur solu-

tions. *Ab initio* molecular dynamics simulations of mABA and $mABA^{\pm}$ in aqueous solution, where meta-aminobenzoic has been reported 50% in the zwitterionic form, were used to characterise the interaction of these molecules with the surrounding water molecules. The cluster-continuum approach[112] was then applied to quantify the equilibrium of mABA and $mABA^{\pm}$ in dimethyl sulfoxide solutions, where the distribution of the zwitterionic and non-zwitterionic forms was still under debate. Microsolvation density functional theory calculations were finally conducted to characterise the energetic barrier for the desolvation of mABA molecules and the strength of mABA-solvent interactions, which were both found to be larger in dimethyl sulfoxide compared with water.

3.1 Introduction

As mentioned in the introduction, the substance meta-aminobenzoic acid is an important active pharmaceutical ingredient and has a very strong polymorphic character[95], which can be related to the manifold of intermolecular interactions between meta-aminobenzoic acid molecules (hydrogen (H) bonding, π-π interactions and H$-\pi$ interactions) but also to the ability of this molecule to exist in either of both the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) forms (Figure 3.1).[113]

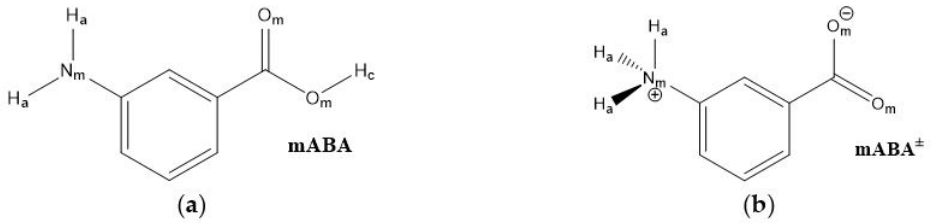

Figure 3.1: Schematic picture of the two tautomeric forms of meta-aminobenzoic acid: (a) nonionic mABA; (b) zwitterionic mABA$^{\pm}$. The oxygen and nitrogen atoms of mABA or mABA$^{\pm}$ are denoted by O_m and N_m, the hydrogen of amino group are denoted by H_a, the hydrogen atoms of carboxylic group are denoted by H_c.

In the polymorphs denoted I, III and IV the molecules of meta-aminobenzoic acid are zwitterionic, and in the polymorphs II and V they are nonionic.[99] The nature of the solvent can significantly influence the thermodynamics and kinetics of crystal growth[71,114] and, consequently, control the formation of one specific polymorph over another.[115,116] In the case of meta-aminobenzoic acid, Form II preferentially crystallises from dimethyl sulfoxide (DMSO),[99] where meta-aminobenzoic acid only exist in the nonionic form. Hughes and co-workers monitored the crystallisation of meta-aminobenzoic acid from organosulfur solutions using a combined liquid- and solid-state *in-situ* NMR apparatus and proposed the existence of nonionic mABA aggregates linked by H bonds;[117] the authors could not, however, uniquely determine the identity of these species. On the other hand, Form I preferentially crystallises from aqueous environments[99] despite it has been reported that the values of the equilibrium constant $K_Z = [mABA^{\pm}]/[\text{mABA}]$ for aminobenzoic acids are of the order of unity in water,[97,98] implying a comparable distribution of mABA$^{\pm}$ and mABA molecules.

As the desolvation of the monomers of meta-aminobenzoic acid represents the first step in the crystallisation process, the intermolecular interactions

between the solute and the surrounding solvent molecules can have a significant influence on the kinetics and thermodynamics of organic crystal formation. This chapter reports a combination of atomistic simulations based on density functional theory (DFT) to study the solvation of the non-ionic and zwitterionic forms of meta-aminobenzoic acid. *Ab initio* molecular dynamics (AIMD) of mABA and $mABA^{\pm}$ in aqueous solution, where meta-aminobenzoic has been reported 50% in the zwitterionic form, were used to characterise the interaction of these molecules with the surrounding water molecules. The cluster-continuum approach,[112] where the solute is comprised of a certain number of solvent molecules treated at the DFT level whereas the bulk solvation environment is described using an implicit model, was applied to quantify the equilibrium of mABA and $mABA^{\pm}$ in dimethyl sulfoxide solutions. Microsolvation density functional theory calculations were finally used to characterise the energetic barrier for the desolvation of mABA molecules and strength of mABA-solvent interactions.

3.2 Methods

3.2.1 Ab initio molecular dynamics simulation

AIMD simulations were conducted with the electronic structure code CP2K, version 4.1. We used the PBE generalised gradient approximation for the exchange and correlation terms together with the general dispersion correction termed DFT-D3. Goedecker-Teter-Hutter pseudopotentials[118] were used to describe the core-valence interactions. All atomic species were rep-

resented using a double-zeta valence polarised basis set. The plane wave kinetic energy cut off was set to 1000 Ry. k-sampling was restricted to the Γ point of the Brillouin zone. Simulations were carried out with a wave function optimisation tolerance of 10^{-6} au that allows for 1.0 fs time steps with reasonable energy conservation. Periodic boundary conditions were applied throughout. Simulations were carried out using one mABA molecule (zwitterionic and non-ionic forms) immersed in 210 water molecules in the NVT ensemble using a Nosé-Hoover chain thermostat to maintain the average temperature at T = 300 K.

3.2.2 Static density functional theory calculations

Electronic structure calculations were carried out with the NWChem (version 6.3) and Gaussian09 codes. We used the Grimme's density functional including dispersion (B97-D2) and the Minnesota 06 global hybrid functional with 54% HF exchange (M06-2X). The Gaussian 6-31+G(d,p) basis set was used throughout these simulations as this provides a good compromise between accuracy and computational cost.[82,119,120] Thermal contributions were calculated at the optimised geometries using the gas-phase harmonic frequencies, which were scaled by a factor of 0.979 to account for systematic errors in the density functional and for anharmonicity.[121] Free energies of solvation were calculated using the SMD solvation model[106] and the gas-phase optimised geometries.

DFT microsolvation modelling. The low-lying energy minima on the PES of the microsolvated zwitterionic, $mABA^{\pm}(S)_n$, and non-zwitterionic, $mABA(S)_n$, forms of m-aminobenzoic acid (S = DMSO or H_2O; n = 1-3),

were located using the following strategy: 1) Several tens of thousands of candidate conformations of the mABA(S)$_n$ and mABA$^{\pm}$(S)$_n$ clusters were generated using the Granada code. 2) The initial set was then reduced with a distance-based selection similar to that used for the generation of mABA molecular clusters. 3) The energy of the candidate structures and associated Boltzmann factors of each configuration were computed at the B97-D2/6-31+G(d,p) level of theory. 4) For the conformations with a Boltzmann factor $f_i \geq 0.01$ the optimised geometries, thermochemical properties and solvation contributions were determined at the M06-2X/6-31+G(d,p) level of theory.

Cluster continuum approach to compute solvation free energies. The SMD continuum model can provide accurate solvation free energies for the non-zwitterionic form of mABA: using the SMD/M06-2X/6-31+G^{**} method the mean unsigned error in the calculated solvation free energies of carboxylic acids was only 0.25 kcal.mol^{-1} in water and 0.55 kcal.mol^{-1} in organic solvents.[106] In contrast, due to strong electrostatic effects arising from unbalanced charges, solvation modeling of the zwitterionic form ($mABA^{\pm}$) is particularly challenging for continuum approaches.[23] To correct for these deficiencies, the solvation free energy of $mABA^{\pm}$ was determined using the cluster-continuum model:[112] the solvated molecule $(mABA^{\pm})(S)_n$ (n = 0 - 3) was treated quantum mechanically (M06-2X/6-31+G^{**}) and the bulk solvent was described with the SMD continuum model. According to this approach the solvation of $mABA^{\pm}$ is given by

$$\begin{aligned}\Delta G^{*}_{solv}(mABA^{\pm}) = \Delta G^{0}_{clust}(mABA^{\pm}(S)_n) \\ +\Delta G^{*}_{solv}(mABA^{\pm}(S)_n) + n\Delta G_{vap}(S)\end{aligned} \tag{3.1}$$

where $\Delta G^{0}_{clust}(mABA^{\pm}(S)_n)$ is the gas-phase free energy of cluster formation at 1 atm, $\Delta G^{*}_{solv}(mABA^{\pm}(S)_n)$ is the solvation free energy of the solvated cluster, and $\Delta G_{vap}(S)$ is the free energy of vaporisation free energy of one solvent molecule. The last term was computed using the following equation:

$$\Delta G_{vap}(S) = -\Delta G^{*}_{solv}(S) - RTln[\tilde{R}T] - RTln[S] \tag{3.2}$$

where $\Delta G^{*}_{solv}(S)$ is the solvation energy of a solvent molecule and the last term is related to the density number of the solvent, which was computed by taking [DMSO] = 14.1 $mol.L^{-1}$ and [H_2O] = 55.5 $mol.L^{-1}$,[27] respectively. For the solvation free energy of DMSO we used the value computed at the SMD/MP2/def2-TZVPP level of theory (-7.75 $kcal.mol^{-1}$).

3.3 Hydration of meta-aminobenzoic acid

This section is concerned with the stability of the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) forms of meta-aminobenzoic acid in aqueous solution, and with the interaction of these molecules with the surrounding water molecules. Hereafter, the oxygen and nitrogen atoms of mABA or $mABA^{\pm}$ are denoted by O_m and N_m, the hydrogens of amino group are denoted by H_a, the hydrogen atoms of carboxylic group are denoted by H_c, and oxygen and hydrogen of water are denoted by O_w and H_w, respectively (Fig. 3.1).

Figure 3.2 reports the time evolution of the intra-molecular (O_m-$H_c$and N_m-H_a) and inter-molecular (O_m...H_w and N_m...H_w) distances during the AIMD simulations of the mABA and $mABA^{\pm}$ species in water. If 1 Å is taken as the average intramolecular X_m-H_w (X = N, O) bond distance, then mABA and $mABA^{\pm}$ are not involved in any proton transfer reactions with the surrounding water molecules. Both mABA and $mABA^{\pm}$ molecules are therefore stable in water and should be considered when modeling the aggregation of meta-aminobenzoic acid in aqueous solution. If we use 2.5 Å to define the existence of intermolecular O_m...H_w and N_m...H_w interactions, then, as the insets of Figure 3.2 (a,b) show, the interaction of mABA with the surrounding water molecules occurs during a very short time range ($<$ 5 ps).

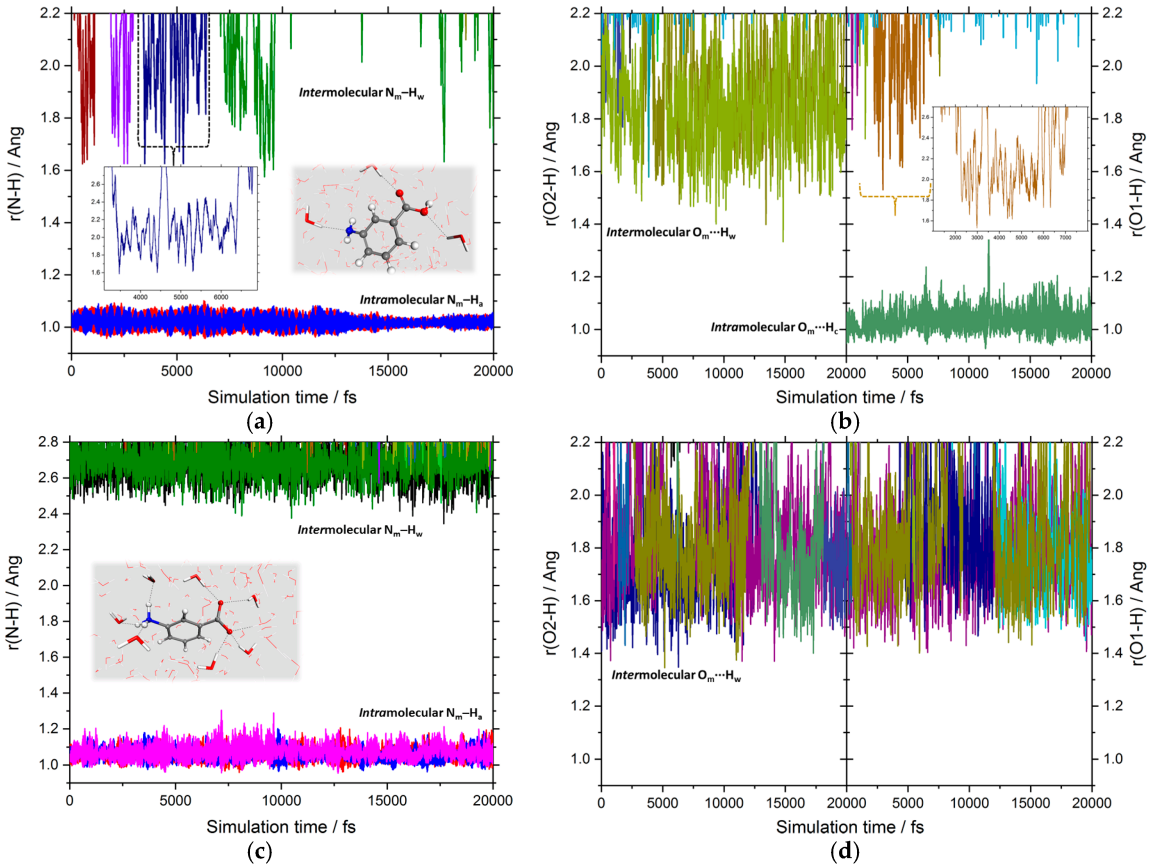

Figure 3.2: Time evolution of the X_m-H (X = N, O) distances during the AIMD simulation of the nonionic (mABA) and the zwitterionic ($mABA^{\pm}$) forms of meta-aminobenzoic acid in water: (a) intramolecular (N_m-H) and intermolecular (N_m...H) distances of the mABA molecule; (b) intramolecular (O_m-H) and intermolecular (O_m...H) distances of the mABA molecule; (c) intramolecular (N_m-H) and intermolecular (N_m...H) distances of the $mABA^{\pm}$ molecule; (d) intramolecular (O_m-H) and intermolecular (O_m...H) distances of the $mABA^{\pm}$ molecule.

A detailed characterisation of intermolecular O_m...H_w and N_m...H_w interactions can be obtained from the analysis of the radial distribution function (RDF), $g_{\alpha\beta}(r)$, which represents the probability relative to a random distribution of finding an atom of type β at a distance r from an atom of type α. Figure 3.3 reports the O_m-H_w and N_m-H_w RDFs together with the running coordination number, $n(r) = (4\pi N/V)\int_0^r g(r')\,dr'$, where N is the

number of hydrogen or oxygen atoms and V is the volume of the simulation cell. In the X_m-H_w (X = N or O) RDFs, a maximum in the [1.5-2.0] Å region and a minimum at around 2.5 Å indicates the presence of an H-bond with the surrounding water molecules.[122] On average, less than one water molecule is coordinated to each oxygen atom of the -COOH group and to the nitrogen atom of the $-NH_2$ group. On the other hand, approximately four water molecules are coordinated to the $-COO^-$ group of $mABA^{\pm}$ and no water molecule is H-bonded to the nitrogen atom of $-NH_3^+$. Table 3.1 summarises the positions (r_{max} and r_{min}) and amplitudes (g_{max} and g_{min}) of the maxima and minima of the X_m-H_w (X = N or O) RDFs together with the ratios $g_{max}^{X_m-H_w}/g_{min}^{X_m-H_w}$; these values can be used as a proxy for the strength of the H-bonding interactions between the X_mH_w pairs (X = O, N).[122,123] For mABA, the $g_{max}^{O_m-H_w}/g_{min}^{O_m-H_w}$ ratio of the carboxyl oxygen atoms (9.0) is higher than that of nitrogen (4.5) but lower than the value of $g_{max}^{O_w-H_w}/g_{min}^{O_w-H_w} = 19.6$ obtained from AIMD simulations of pure water. Similar behavior is observed for $mABA^{\pm}$, but the interaction of the COO^- group ($g_{max}^{O_w-H_w}/g_{min}^{O_w-H_w} = 14.0$) is significantly stronger than mABA.

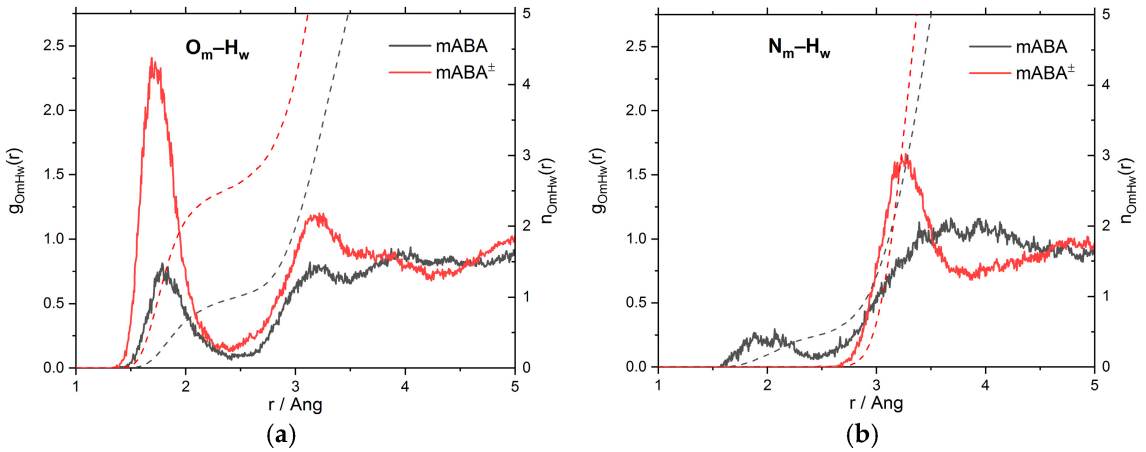

Figure 3.3: The radial distribution functions, g(r) (continuous lines), and running coordination numbers, n(r) (dashed lines), of mABA and mABA$^{\pm}$ with water obtained from AIMD simulations: (a) O_m-H_w RDFs (O_m = oxygen atoms of meta-aminobenzoic acid; H_w = hydrogen atoms of water); (b) N_m-H_w RDFs (N_m = nitrogen atoms of meta-aminobenzoic acid; H_w = oxygen atoms of water). The integration number has been computed as $n(r) = \int 4\pi N/V \int_0^r g(r)dr$, where N is the number of hydrogen atoms and V is the volume of the simulation cell.

Table 3.1: Positions (r_{max}^{X-H} and r_{min}^{X-H} in Å) and amplitudes (g_{max}^{X-H} and g_{min}^{X-H}) of the maxima and minima of the first peak of the $X_m - H_w$ (X = O_m, N_m) RDFs, and first hydration shell numbers (n_w) obtained from the AIMD simulations of mABA and mABA$^{\pm}$ in water.

	mABA	mABA$^{\pm}$
$r_{max}^{O_m-H_w}$	1.79	1.72
$g_{max}^{O_m-H_w}$	0.81	2.38
$r_{min}^{O_m-H_w}$	2.50	2.52
$g_{min}^{O_m-H_w}$	0.09	0.17
$g_{max}^{O_m-H_w}/g_{min}^{O_m-H_w}$	9.00	14.00
$n_w^{O_m}$	1.0	2.6
$r_{max}^{N_m-H_w}$	1.88	-
$g_{max}^{N_m-H_w}$	0.27	-
$r_{min}^{N_m-H_w}$	2.46	-
$g_{min}^{N_m-H_w}$	0.06	-
$g_{max}^{N_m-H_w}/g_{min}^{N_m-H_w}$	4.50	-
$n_w^{N_m}$	0.5	0

The RDFs and structural data of the H_c-O_w and H_a-O_w intermolecular interactions are reported in Figure 3.4 and Table 3.2. For the carboxylic group of mABA, the H_c-O_w RDF has a very well-defined maximum at 1.51 Å, and the running coordination number ($n_w{}^{H_c}$) is characterised by a clear plateau at the first RDF minimum (Figure 3.4a). The value of $g_{max}^{H_c-O_w}/g_{min}^{H_c-O_w}$ is significantly larger than $g_{max}^{O_w-H_w}/g_{min}^{O_w-H_w}$ of pure water (19.6), so the H_c-O_w interaction is stronger than the intermolecular H-bonding in bulk water. The hydrogen of -COOH is therefore stably coordinated to a single water molecule. For the amino group of mABA, the

hydrogen atoms of the -NH_2 group do not interact significantly with the surrounding water molecules because, in the [1.5-2.0] Å region, the H_a-O_w RDF is not characterised by a well-defined peak (Figure 3.4b). On the other hand, the H_a-O_w RDF of the -NH_3^+ in mABA$^{\pm}$ is characterised by a distinct peak at 1.77 Å.

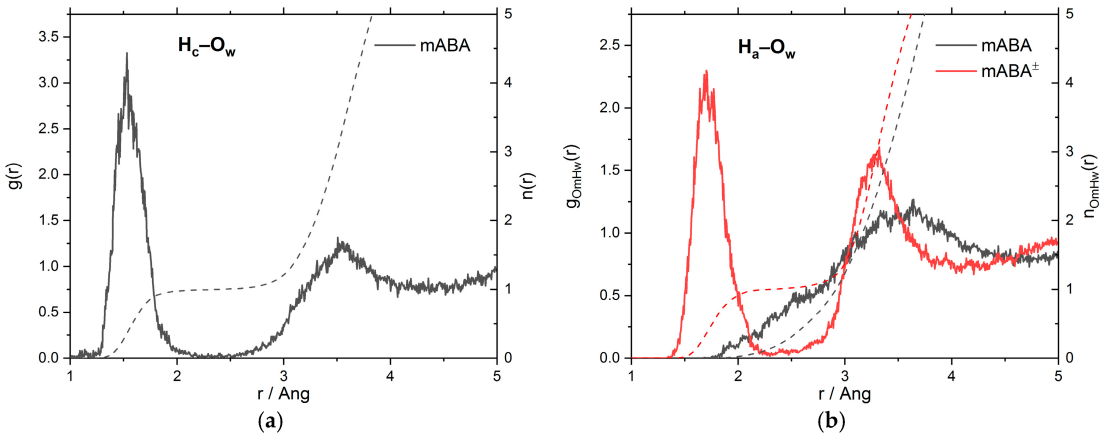

Figure 3.4: The radial distribution functions, g(r) (continuous lines), and running coordination numbers, n(r) (dashed lines), of mABA and mABA$^{\pm}$ with water obtained from AIMD simulations: (a) $H_c - O_w$ RDFs (O_c = oxygen atoms of the carboxylic group of mABA; H_w = hydrogen atoms of water); (b) $H_a - O_w$ RDFs N_m = nitrogen atoms of the amino group of mABA and mABA$^{\pm}$; O_w = oxygen atoms of water). The integration number has been computed as $n(r) = \int 4\pi N/V \int_0^r g(r)dr$, where N is the number of hydrogen atoms and V is the volume of the simulation cell.

Table 3.2: Positions (r_{max}^{H-O} and r_{min}^{H-O} in Å) and amplitudes (g_{max}^{H-O} and g_{min}^{H-O}) of the maxima and minima of the first peak of the $H_a - O_w$ and $H_c - O_w$ RDFs, and first hydration shell numbers (n_w) obtained from the AIMD simulations of mABA and mABA$^{\pm}$ in water.

	mABA	mABA$^{\pm}$
$r_{max}^{H_c-O_w}$	1.51	-
$g_{max}^{H_c-O_w}$	3.06	-
$r_{min}^{H_c-O_w}$	2.31	-
$g_{min}^{H_c-O_w}$	0.01	-
$g_{max}^{H_c-O_w}/g_{min}^{H_c-O_w}$	306.00	-
$n_w^{H_c}$	1.0	-
$r_{max}^{H_a-O_w}$	-	1.77
$g_{max}^{H_a-O_w}$	-	2.15
$r_{min}^{H_a-O_w}$	-	2.23
$g_{min}^{H_a-O_w}$	-	0.03
$g_{max}^{H_a-O_w}/g_{min}^{H_a-O_w}$	-	71.7
$n_w^{H_a}$	0	1.0

The analysis of the X_m-H_w (X = N, O), H_c-O_w and H_a-O_w RDFs indicates therefore that in aqueous solution the mABA$^{\pm}$-water interaction is mostly stronger than mABA-water. The value between the maximum and minimum of the RDFs shows that the (X_m-H_w) interaction is stronger for mABA$^{\pm}$ than for mABA. Moreover, the interaction of both species with the surrounding water molecules is stronger around the carboxylic acid than around the amino group.

The probability distribution of the number of water molecules in the first hydration shell (HS) of mABA and mABA$^{\pm}$ was determined from the pair

correlation functions between the center-of-mass (COM) of meta-aminobenzoic acid and the COM of the water molecules (Figure 3.5). The position of the first HS was approximated by the first minimum in the COM(mABA)-COM(H_2O) RDFs (insets of Figure 3.5), and although a hydration shell can be located for both molecules, the probability distributions of the number of water molecules surrounding mABA and mABA$^{\pm}$ show the flexibility of their HS: there are, on average, 24 water molecules in the HS of mABA with a mean absolute deviation (MAD) of 1.4, and 27 water molecules in the HS of mABA$^{\pm}$ with an MAD of 1.0.

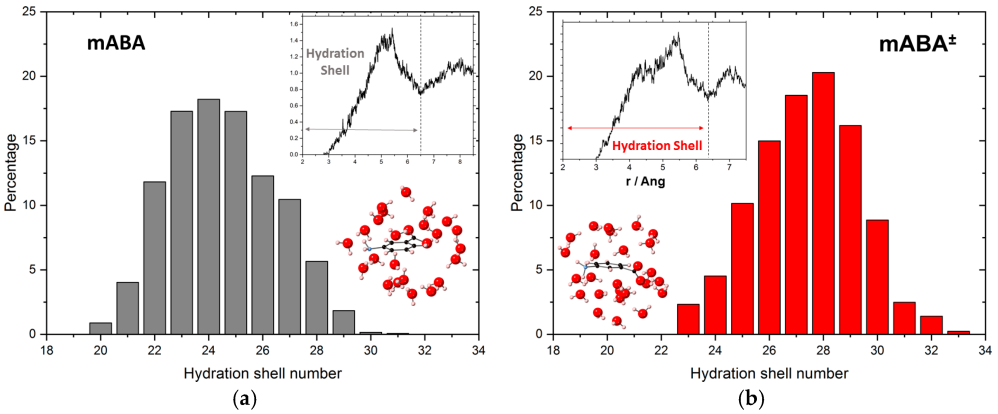

Figure 3.5: (a) Probability distribution of the coordination number in the hydration shell of mABA, the mABA-H_2O radial distribution function of the center-of-masses of mABA and water (inset), and the optimised structure of mABA with its hydration shell. (b) Probability distribution of the coordination number in the hydration shell of mABA$^{\pm}$, the mABA$^{\pm}$-H_2O radial distribution function of the center-of-masses of mABA$^{\pm}$ and water, and the optimised structure of mABA$^{\pm}$ with its hydration shell.

3.4 Equilibrium of zwitterionic non-zwitterionic forms in DMSO

To interpret measurements of solid- and liquid-NMR spectra of mABA, it was hypothesised that in dimethyl sulfoxide solutions, molecules of mABA are predominantly non-zwitterionic.[117] This is not an obvious conclusion because the polarity index (P') of DMSO (P' = 7.2) is higher than most other solvents, including polar protic solvents such as methanol (P' = 5.1).[124] In order to verify this hypothesis and determine which form of mABA should be modelled in this study, the microscopic constants pertaining to the zwitterionic equilibrium was calculated according to the following expression:

$$K_Z = \frac{mABA^{\pm}}{mABA} = e^{-\frac{\Delta G_Z^*}{RT}} \tag{3.3}$$

where $\Delta G_Z{}^*$ is the free energy for the zwitterionic equilibrium mABA $\rightleftharpoons$ mABA$^{\pm}$ in DMSO and is given by:

$$\Delta G_Z^* = \Delta E_{e,gas} + \Delta\delta G^0_{VRT,gas} + \Delta\Delta G^0_{solv} \tag{3.4}$$

Reliable estimates of the free energy of this reaction depend on the accurate determination of each term in equation 3.4, but when the species involved in the equilibrium are ionic then the accurate evaluation of the solvation free energy term ($\Delta\Delta G_{solv}{}^0$) becomes particularly critical.[112] Table 3.3 reports the solvation free energies of mABA$^{\pm}$ computed according the cluster continuum approach (see Eq. 3.1) using up to three explicit DMSO solvent molecules in the microsolvated mABA$^{\pm}$(S)$_n$ model. The

most stable structures of the $mABA^{\pm}(DMSO)_n$ clusters are displayed in Figure 3.6.

Table 3.3: Solvation free energies in DMSO of the zwitterionic ($mABA^{\pm}$) form of m-aminobenzoic acid obtained using the cluster-continuum approach. Gas-phase free energy contributions computed at the M06-2X/6-31+G(d,p) level and solvation free energies computed at the SMD/M06-2X/6-31+G(d,p) level. Values in kcal.mol^{-1}

	$\Delta G_{clust}{}^{0}(mABA^{\pm})(S)_n$	$\Delta G_{solv}{}^{*}(mABA^{\pm})(S)_n$	$\Delta G_{solv}{}^{*}(mABA^{\pm})$
0	-	-	-56.55
1	-26.16	-19.34	-59.03
2	-39.53	-37.15	-62.20
3	-50.72	-31.25	-62.72

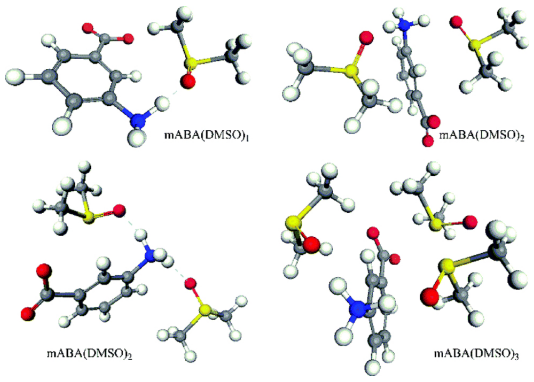

Figure 3.6: Optimised gas phase structures of the most stable solvated $mABA^{\pm}(DMSO)_n$ (n = 1-3) clusters as computed at the M06-2X/6-31+G(d,p).

According to Pliego and Riveros,[125] a variational principle can be established in the cluster continuum approach for the choice of the number of explicit solvent molecules n: the value that minimises $\Delta G_{solv}{}^{*}$ is the ideal number of explicit solvent molecules and the solvation free energy for this value is the best calculated value for $\Delta G_{solv}{}^{*}$. For $mABA^{\pm}(S)_n$ this

corresponds to $n = 3$ and the best estimate of the free energy of solvation for $mABA^{\pm}$ is -62.72 kcal.mol^{-1}. This high value for the solvation free energy of $mABA^{\pm}$ is related to the localisation of the negative and positive charges in the carboxylic (-COO^-) and amino (-NH_3^+) groups, which strongly interact with the polar DMSO solvent molecules. In comparison, the experimental hydration free energies of the $HCOO^-$ and $CH_3NH_3^+$ ions are -74.6 kcal.mol^{-1} and -75.2 kcal.mol^{-1}, respectively.[125]

The energetics of the zwitterionic equilibrium mABA $\rightleftharpoons$ $mABA^{\pm}$ (see Eq. 3.4) are listed in Table 3.4. The solvation term $\Delta\Delta G_{solv}{}^*$ is given by the difference between the solvation free energies of $mABA^{\pm}$ (-62.72 kcal.mol^{-1}) and mABA (-11.29 kcal.mol^{-1}, SMD/M06-2X/6-31+G(d,p) level of theory), whereas the gas-phase energy contribution was computed using different density functionals (PBE, B3LYP, M06-2X) and at the *ab initio* MP2 level. All methods agree that the free energy of the zwitterionic equilibrium mABA $\rightleftharpoons$ $mABA^{\pm}$ is positive (between 7.2 and 9.5 kcal.mol^{-1}), supporting the hypothesis made by Hughes and co-workers that in DMSO solutions, molecules of mABA principally exist in the non-zwitterionic form.[117]

Table 3.4: Energetics of the zwitterionic equilibrium mABA $\rightleftharpoons$ mABA$^{\pm}$. Gas-phase energies ($\Delta E_{e,gas}$) and standard state (1 atm) gas-phase free energies computed with different quantum chemistry methods. Geometries, frequencies and solvation energies computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal.mol^{-1}

Method	$\Delta E_{e,gas}$	$\Delta G_{z,gas}{}^{0}$	$\Delta\Delta G_{solv}{}^{*}$	$\Delta G_{z,solv}{}^{*}$
M06-2X/6-31+G(d,p)	61.78	60.90	-51.43	9.47
M06-2X/aug-cc-pVTZ	61.57	60.69	-51.43	9.26
PBE/aug-cc-pVTZ	57.91	57.03	-51.43	5.60
B3LYP/aug-cc-pVTZ	60.00	59.12	-51.43	7.69
MP2/aug-cc-pVTZ	59.55	58.67	-51.43	7.24

3.5 Quantification of the strength of mABA-solvent interaction

Solute-solvent interaction can have an important effect on the kinetics of formation of organic crystals. Khamar *et al.* concluded that the influence of the solvent on the crystal nucleation is related to the process of salicylic acid desolvation: the stronger the interaction between the solvent molecules and the salicylic acid molecule, the slower the crystal nucleation in solution.[32] Here, the energy (ΔE_{clust}) and free energy ($\Delta G_{clust}{}^{0}$) of the solute-solvent clustering process and the successive solvent binding energy (Δ E_{inc}) are used as descriptors of the mABA-solvent interaction:

$$\Delta E_{clust} = E_{[mABA(S)_n]} - (E_{[mABA]} + nE_{[S]}) \tag{3.5}$$

$$\Delta G_{clust} = G^{0}_{[mABA(S)_n]} - (G^{0}_{[mABA]} + nG^{0}_{[S]}) \tag{3.6}$$

$$\Delta E_{inc} = E_{[mABA(S)_n]} - E_{[mABA(S)_{n-1}]} - E_{[S]} \tag{3.7}$$

In equations 3.5 to 3.7, E and G^0 are the gas-phase energies and gas-phase free energies of the solvated cluster [mABA$(S)_n$], solute (mABA) and solvent (H_2O or DMSO). These values were computed at the M06-2X/6-31+G(d,p) level of theory. The structures of the low-lying structures of isomers of mABA$(S)_n$ (n = 1-3) were determined using the computational protocol outlined in the computational section starting from tens of thousands of candidate structures. The energetics of binding and successive binding are reported in Table 3.5. For comparison, the solvation free energies of mABA in water and DMSO computed at the SMD/M06-2X/6-31+G(d,p) level of theory are -10.97 and -11.29 kcal.mol^{-1}, respectively.

Table 3.5: Solute-solvent binding energy (ΔE_{clust}), free energy of solute-solvent clustering (ΔG_{clust}) and successive solvent binding energy (ΔE_{inc}) and solvation free energies of m-aminobenzoic acid in water and in DMSO. Geometries, frequencies and solvation energies computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal.mol^{-1}

	ΔE_{clust}		ΔG_{clust}		ΔE_{inc}	
Species	H_20	DMSO	H_20	DMSO	H_20	DMSO
mABA$(S)_1$	-7.88	-17.68	-7.88	-17.68	2.40	-5.80
mABA$(S)_2$	-22.43	-30.21	-14.55	-12.52	-0.43	-4.90
mABA$(S)_3$	-33.06	-43.20	-10.63	-12.99	0.34	-5.49

The energetics in Table 3.5 show that the clustering energies and free energies of mABA with DMSO are substantially larger (up to 10 kcal.mol^1) than with water. The energy required to desolvate mABA, which is related with the specific strength of the solute-solvent interaction, is therefore a crucial factor in determining the level of clustering, and subsequent growth,

of mABA in water and in DMSO. These topics which are discussed in the following chapters of this thesis.

3.6 Conclusions

The solvation of meta-aminobenzoic acid in water and DMSO were investigated by means of atomistic simulations. *Ab initio* molecular dynamics (AIMD) of two aqueous solutions containing one mABA and one $mABA^{\pm}$ molecules in approximately 200 water molecules were conducted to determine the stability, intermolecular and hydration properties of these two species. A detailed analysis of the number and strength of hydrogen bonds of mABA and $mABA^{\pm}$ with the surrounding water molecules shows that the $mABA^{\pm}$-water interaction is stronger than mABA-water, and that the interaction with the surrounding water molecules is stronger around the carboxylic acid than around the -NH_2 (mABA) and -NH_3^+ ($mABA^{\pm}$) groups. Although a coordination shell can be located for both molecules, the probability distributions of the number of water molecules surrounding mABA and $mABA^{\pm}$ show a great degree of flexibility of the hydration environment.

A cluster-continuum approach was used to compute the energetics of the zwitterionic equilibrium mABA $\leftrightarrow$ $mABA^{\pm}$ and determine the distribution of the zwitterionic and nonionic species in DMSO. Our calculations confirm that the free energy of the zwitterionic equilibrium mABA $\leftrightarrow$ $mABA^{\pm}$ is highly positive (between 7.2 and 9.5 $kcal.mol^{-1}$), supporting the hypothesis made by Hughes and co-workers that in DMSO solutions, molecules of mABA principally exist in the non-zwitterionic form.

DFT calculations of the energy and free energy of formation of microsolvated mABA clusters, $mABA(S)_n$, show that mABA-DMSO interaction is substantially larger (up to 10 $kcal.mol^{-1}$) than mABA-water interaction. During the crystallisation of inorganic crystals such as magnesite ($MgCO_3$) or calcite ($CaCO_3$), the desolvation of the building units, and particularly of the cations,[23,27] can be rate-determining. The computational results obtained in this study suggest that the early stages of crystallisation of organic crystals can be significantly influenced by the desolvation of mABA molecules in solution: the solvent and its specific interaction with the organic solute molecules can influence not only the thermodynamics of aggregation but also the rate of desolvation in DMSO compared with water, which limits the extend of mABA self-assembly process in DMSO. This result could have implications for the computational methods that can be employed to model the crystallisation of organic molecules from solution: approaches based on a continuum description of the solvent to describe the thermodynamic stability of the molecular aggregation,[126] or iterative methods where solvent molecules are excluded during the aggregation step,[127] cannot quantify the desolvation of molecules and could therefore overlook rate-determining processes such as desolvation of the building units in homogeneous nucleation or the desolvation of the surface during surface growth.

Chapter 4

The role of solvent in the self-assembly of m-aminobenzoic acid

Solvent can have significant effects on the solution thermodynamics and crystallisation kinetics of organic compounds from solution. In this chapter, the early stages of aggregation of the organic molecule mABA in two different solvents, DMSO and water, were studied using a combination of quantum chemistry, molecular dynamics and metadynamics simulations. Density functional theory (B97-D and M06-2X) calculations with the continuum solvation SMD model were used to probe the potential energy surface of molecular clusters $(\mathrm{mABA})_n$ (n = 2-4), locate the low-lying energy structures, and compute the Gibbs free energies of $(\mathrm{mABA})_\mathrm{n}$ in solution. In DMSO, the formation of the classic carboxylic $(\mathrm{mABA})_2$ dimer is exergonic and thermodynamically more favourable than in water. Molecular dynamics simulations of mABA solutions at different concentrations show a sig-

nificant solvent-dependent aggregation behaviour of mABA: in water, even at low concentrations, mABA molecules spontaneously form H-bonded π-π stacking molecular clusters whereas in organosulfur solutions the molecules of mABA are in a much more solvated state. Metadynamics simulations are used to rationalise the low level of mABA aggregation in DMSO, compared with water, in terms of the free energy barrier for the desolvation of mABA molecules and formation of dimers. This chapter shows therefore how the solvent and its specific interactions with the organic solute molecules influences both the thermodynamics and kinetics of molecular self-assembly.

4.1 Introduction

In the preparation of pharmaceutical materials that require a well-defined crystal form, the phenomenon of polymorphism can be a threat. Polymorphs can have significantly different solid-state properties and therefore different performances in materials applications, or bioavailability and stability as a drug substance.[128,129] Consequently, one of the major challenges in the synthesis of pharmaceuticals is the selection of a molecular crystal polymorph during solution crystallisation.[130]

It is well known from experimental evidence that the nature of the solvent, or the presence in solution of additives and impurities, can determine the formation of one specific polymorph over another.[131] Several studies have suggested that during the first stages of crystallisation (clustering and nucleation) the solute-solvent interactions can control the polymorphic outcome.[75,132] As the process of molecular self-assembly to form the

crystallisation growth unit is significantly influenced by the interaction between the solute and solvent molecules, namely by the properties of the solvation environment, there is considerable interest, both fundamental and technological, in understanding how the processes of clustering, nucleation and subsequent growth of a molecular crystal are affected by the solution chemistry.[105,133] However, advanced spectroscopic techniques such as synchrotron radiation X-ray scattering or atomic force microscopy cannot give a complete understanding of these processes.[134] The interpretation of experimental measurements and, most importantly, the ability to predict how a new crystal phase nucleates and then grows increasingly depends on the use of computer simulation.[44]

Significant work has been done in the development and application of simulation methods to model the nucleation and growth from solution.[39,76,87,90,92] A recent theoretical investigation has revealed a direct relation between the computed strength of the solute-solvent interaction and the experimental rate of nucleation of organic molecules.[32] This study, however, was based on static density functional theory (DFT) gas-phase calculations of microsolvated organic molecules, which do not fully account for the thermodynamic and kinetic factors affecting the aggregation of organic molecules from solution.

To obtain a better understanding of the role of solvent in the kinetics and thermodynamics of the early stages crystallisation from solution, the aggregation of meta-aminobenzoic acid (mABA) in two different solvents, dimethyl sulfoxide (DMSO) and water, were carried out using a combination of computational methods based on quantum chemistry solvation continuum, molecular dynamics and free energy techniques.

The molecule mABA is a fascinating model for polymorphic research due to its ability to crystallise in five different crystal forms, whose nucleation depends chiefly on the solvent.[93,115,135] The clustering of mABA molecules is also of interest because it can form complex mABA/mABA intermolecular interactions, *i.e.* XH/O and XH/π (X = O, N) hydrogen bonds and $\pi - \pi$ interaction, all of which are important generally during polymorph selection.

Having introduced in Section 4.2 the details of the simulations, we next report DFT calculations of the thermodynamics of formation of the oligomers of m-aminobenzoic acid $(\text{mABA})_n$ (n = 2-4). Results from molecular dynamics simulations of the aggregation of mABA in water and in DMSO are then discussed. The solvent-dependent aggregation behavior of mABA observed from these simulations is rationalised in terms of the free energy profile for mABA dimerisation in water and DMSO.

4.2 Methods

4.2.1 Density functional theory calculations

Electronic structure calculations were carried out with the NWChem (version 6.3)[136] and Gaussian09[137] codes. We used the Grimme's density functional including dispersion (B97-D)[104] and the Minnesota 06 global hybrid functional with 54% HF exchange (M06-2X).[119] The Gaussian 6-31+G(d,p) basis set was used throughout these simulations as this provides a good compromise between accuracy and computational cost.[82,119,120] Thermal contributions were calculated at the optimised geometries using the gas-phase harmonic frequencies, which were scaled by a factor of 0.979 in order

to account for systematic errors in the density functional and for anharmonicity.[121] Free energies of solvation were calculated using the SMD solvation model,[106] and the gas-phase optimised geometries.

Free energies of association in solution. The free energy of formation of carboxylic acid clusters were computed according to the following equation:

$$\Delta G^*_{ass} = G^*_{AB} - G^*_A - G^*_B \tag{4.1}$$

where $G_X{}^*$ is the total Gibbs free energy of the species X (X = AB, A or B) in the liquid at 298 K, which was determined by the addition of the gas-phase total electronic energy of the solute ($E_{e,gas}$), the vibrational-rotational-translational contribution to the gas-phase Gibbs free energy ($\delta G_{VRT,gas}{}^0$) at T = 298 K under a standard-state partial pressure of 1 atm, the solvation free energy of the solute corresponding to transfer from an ideal gas at a concentration of 1 mol.L^{-1} to an ideal solution at a liquid-phase concentration of 1 mol.L^{-1} ($\Delta G_{solv}{}^*$), and the free energy change of 1 mol of an ideal gas from 1 atm to 1 mol.L^{-1} (RTln[$\tilde{R}$T] = 1.89 kcal.mol^{-1} at 298.17 K where $\tilde{R}$T = 0.082052 K^{-1}):

$$G^*_X = E_{e,gas} + \delta G^*_{VRT,gas} + \Delta G^*_{solv} + RTln[\tilde{R}T] \tag{4.2}$$

For multiple stationary points of a molecular cluster, the free energy was determined from the Boltzmann ensemble average

$$\langle G(X) \rangle = \sum_{i=1}^{N} f_i G(X_i) \tag{4.3}$$

where f_i is the Boltzmann factor corresponding to the ith configuration,

$G(X_i)$ is the corresponding free energy and N is the number of low-lying energy structures located on the potential energy surface (PES) of the molecular cluster. The Boltzmann factor was determined according to 4.4.

$$f_i = \frac{e^{-G(X_i)/RT}}{\sum_j e^{-G(X_j)/RT}} \tag{4.4}$$

where R is the universal gas constant, T is the absolute temperature ($T = 298$ K) and the index j runs over all isomers.

To identify the most stable isomers of the molecular clusters $(mABA)_n$ (n = 2-4) we have adopted the following protocol:

- For each m-aminobenzoic cluster $(mABA)_n$, several hundreds of thousands of candidate structures were generated using Granada, a code designed to randomly distribute one or more molecules around a central unit (a monomer, dimer, trimer etc.) placed at the centre of a cube of defined side length. For example, we generated candidate structures for the trimer $(mABA)_3$ by considering the random distribution of one mABA molecule around the most stable dimer (the cyclic dimeric structure), and from the random distribution of two mABA molecules around a central mABA molecule. The molecular units used to generate the candidate structures were fully optimised at the M06-2X/6-31+G(d,p) level of theory.

- Using the configuration selector code, only those configurations such that at least one atom of each mobile molecule was within 4 Å from at least one atom of the central unit were selected as candidate low-lying

energy structures.

- The energies of these structures were evaluated at the B97-D/6-31+G(d,p) level of theory and the Boltzmann factor f_i corresponding to the i^{th} configuration was determined as

$$f_i = \frac{e^{-(E_i - E_0)/RT}}{\sum_j e^{-(E_j - E_0)/RT}} \tag{4.5}$$

  where E_i was the energy of the i^{th} candidate structure and E_0 was the energy of the most stable candidate structure. We then selected the candidate structures with a Boltzmann factor $f_i \geq 0.01$ and to increase our sampling we also selected ten to fifteen randomly selected structures such that $3 \leq E_i - E_0 \leq 15$ kcal.mol^{-1}.

- The full optimised geometries, thermochemical properties and solvation energies of the configurations selected at point (3) were computed at the M06-2X/6-31+G(d,p) level of theory.

We chose M06-2X because its assessment against representative databases showed that this method is one of the most accurate density functionals for applications involving a combination of main-group thermochemistry and noncovalent interactions.[138,139] Moreover, the application of the SMD solvation model with M06-2X to predict the free energies of aqueous solvation for 61 drug-like molecules in the SAMPL1 test set gave a mean unsigned error of only 2.0 kcal.mol^{-1}.[107] However, M06-2X has shown small energy oscillations at distances away from the equilibrium.[140] For this reason, in the protocol discussed above, the energies of the candidate structures were evaluated using the B97-D functional, but the equilibrium geometries, thermochemical properties and solvation energies of the molecular clusters were

computed with the M06-2X method.

4.2.2 Classical molecular dynamics

Forcefield. The general AMBER forcefield (GAFF)[141] was used to model the mABA and DMSO molecules. The GAFF potential is a well established potential to simulate organic molecules and it was previously used to compute processes of aggregation and crystal growth of other organic systems,[87,90,92] including para-aminobenzoic acid.[105] Water molecules were modelled using the SPC/E potential.[142] The interactions between mABA and DMSO molecules and between mABA and water molecules were described using the GAFF potential. To derive the forcefield parameters within the framework of the GAFF, the structures and molecular electrostatic potential (MEP) of mABA and DMSO were obtained using the Gaussian09 code. The Antechamber package was then used to compute partial charges according to the restrained electrostatic potential (RESP) formalism.[143] The selection of the quantum chemistry method and basis set to compute MEP is a key aspect of the RESP derivation.[144] Here, we used the HF/6-31G* method, which was the level of theory applied for the derivation of RESP charges in the Cornell *et al.* forcefield,[145] and successive modifications of the AMBER potential.[141,146–148] Using this procedure, we obtained partial charges for DMSO consistent with the values reported by Dupradeau *et al.*[144] The GAFF forcefields and partial charges used to model mABA and DMSO are reported in the Appendix A.2.

Simulation details. Classical molecular dynamics (MD) simulations were performed using version 5.0.4 of the GROMACS molecular dynamics package.[149,150] The leapfrog algorithm with a time step of 2 fs was used to in-

tegrate the equations of motion. The isothermal-isobaric (constant NPT) ensemble was used to maintain a temperature of 300 K and a pressure of 1 bar. The velocity rescale thermostat and the isotropic Parrinello-Rahman barostat were used with 0.4 ps and 2.0 ps as the thermostat and barostat relaxation times, respectively. The electrostatic forces were calculated by means of the particle-mesh Edwald approach with a cutoff of 1.2 nm. A 1.2 nm cutoff was also used for the van der Waals forces. The LINCS algorithm was applied at each step to preserve the bond lengths.

Simulation protocol. We performed MD simulations of mABA solutions in DMSO and water at different concentrations as reported in table 4.1. Molecular models of mABA solutions were generated using the *insert-molecules* and *solvate* GROMACS utilities to insert the required number of mABA molecules in an empty cubic box of size 5 nm, and solvate them with DMSO or water. Each solution was at first minimised using the conjugate-gradient algorithm with a tolerance on the maximum force of 200 $kJ.mol^{-1}$, and the temperature and volume of each system were equilibrated by running 100 ps of constant volume, constant temperature (NVT) simulation followed by 1 ns of NPT simulations. Production runs in the NPT ensemble were then conducted for tens to hundreds of ns. Details of the simulation times, number of solute and solvent molecules, and equilibrated values of the average cell length are reported in table 4.1.

Metadynamics simulations. The free energy profile for the formation of a carboxylic acid dimer (product) from two fully solvated mABA molecules (reactants) was computed by means of the well-tempered metadynamics method.[151,152] We generated solutions containing two mABA molecules in 945 DMSO and 4058 water molecules, respectively. The free energy calcula-

tions were performed with GROMACS 5.0.4 equipped with the PLUMED 2.2b plugin.[109] Simulations were conducted in the NPT ensemble for 10 ns. The order parameter (OP) used to study the dimerisation reaction was defined as follows:

$$OP = \frac{d_1 + d_2}{2} \tag{4.6}$$

where d_1 and d_2 are the two distances between the hydroxyl hydrogen (O-H) and the carbonyl oxygen (C=O) of a pair of mABA molecules.

Table 4.1: Details of classical molecular dynamics simulation of mABA solutions

	solvent	no. of mABA molecules	no. of solvent molecules	Box length (Å)	[mABA(mol/L)]	simulation time (ns)
1	DMSO	8	3003	70.66	0.04	200
2	DMSO	16	3002	70.74	0.08	200
3	DMSO	32	3036	70.84	0.15	200
4	DMSO	64	3004	71.22	0.29	200
5	DMSO	128	3012	71.92	0.57	200
6	DMSO	256	3030	73.23	1.08	200
7	water	8	11480	70.25	0.04	30
8	water	16	11426	70.13	0.08	30
9	water	32	11314	69.95	0.16	30
10	water	64	11419	70.70	0.30	30
11	water	128	11430	71.27	0.59	30
12	water	256	11075	72.03	1.14	30

4.3 Role of solvent in the energetics of cluster formation

Table 4.2 reports the energies and free energies for the formation, in the gas phase and in solution, of mABA clusters consisting of two (dimer),

three (trimer) and four (tetramer) molecular units. Molecules of mABA can interact via H-bonding, π-π interactions, H-π interactions (via the hydroxy- or amino- group). Consequently, for each $(mABA)_n$, there can be several low-lying energy isomers, which would be difficult to locate only by means of chemical intuition. For each of the $(mABA)_n$ cluster, the values are obtained by computing the Boltzmann average of the energies (or free energies) of the top stable isomers located using the computational protocol described in section 4.2.

Table 4.2: Energetics of formation of meta-aminobenzoic acid clusters in the gas-phase and solution as computed at the M06-2X/6-31+G(d,p) level of theory. Values obtained from the Boltzmann average of the energies or free energies of the low-lying $(mABA)_n$ isomers. Values in kcal.mol^{-1}

Species	Reaction	$\Delta E_{e,gas}$	$\Delta G_{ass}{}^{0}$	$\Delta G_{ass}{}^{*}$	
				H_2O	DMSO
Dimer	mABA + mABA $\rightarrow$ $(mABA)_2$	-18.33	-6.64	-0.07	-3.40
Trimer	$mABA_2$ + mABA $\rightarrow$ $(mABA)_3$[a]	-13.59	2.12	1.22	5.34
	3 mABA $\rightarrow$ $(mABA)_3$[a]	-31.95	-5.03	0.98	1.29
	3 mABA $\rightarrow$ $(mABA)_3$[b]	-19.98	5.08	7.68	7.96
Tetramer	$(mABA)_2$ + $(mABA)_2$ $\rightarrow$ $(mABA)_4$[c]	-20.77	-0.78	-2.78	2.06
	$(mABA)_2$ + 2 mABA $\rightarrow$ $(mABA)_4$[c]	-39.09	-7.41	-2.99	-1.36

[a] $(mABA)_3$ isomers generated from the condensation of the cyclic dimer I with a mABA molecule.
[b] $(mABA)_3$ isomers generated from the condensation of three isolated mABA molecules. [c] $(mABA)_4$ isomers generated from the condensation of two cyclic dimers.

4.3.1 Dimers

Dimers are the first species that could form in solution during the molecular aggregation process, and there has been much debate on the link between the most populated dimers in solution and the structural motifs in

polymorphs.[75,76,82,153] We determined the structures and free energies of 32 stable $(mABA)_2$ structures, which were classified into 12 possible isomers on the PES of $(mABA)_2$ (see Fig. 4.1). The energetics of formation of each type of dimer are reported in table 4.3.

Table 4.3: Energetics of formation of m-aminobenzoic acid dimers in the gas-phase and solution as computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal.mol^{-1}.

Species	Reaction	Type	$\Delta E_{e,gas}$	ΔG_{ass}	$\Delta G_{ass}{}^{*}$	
					H_20	DMSO
dimer	mABA + mABA $\rightarrow$ $(mABA)_2$	I	-18.33	-6.64	-0.20	-3.40
		II	-13.77	1.41	6.06	5.30
		III	-12.03	-0.22	3.36	2.05
		IV	-11.67	1.95	2.70	4.00
		V	-11.59	-0.44	4.35	3.49
		VI	-10.19	1.77	2.97	4.74
		VII	-9.49	2.60	0.98	3.40
		VIII	-9.31	1.14	4.56	3.07
		IX	-8.90	3.09	4.17	6.83
		X	-8.83	3.67	3.68	6.45
		XI	-8.27	3.29	4.65	4.94
		XII	-4.88	4.83	5.18	5.77

The classic carboxylic acid dimer I is the most stable dimer in both the gas-phase and solution. In particular, the formation of the cyclic dimeric structure I is highly exergonic in DMSO ($\Delta G_{ass}{}^{*}$ = -3.40 kcal.mol^{-1}) compared with water ($\Delta G_{ass}{}^{*}$ = -0.20 kcal.mol^{-1}). For all other dimers (II-XII in Fig. 4.1), where the mABA molecules interact via other types of H-bonding

(*e.g.* NH_2...OH) or weaker π-π, OH/π and NH_2/π interactions, the energetics of formation is positive in both water and DMSO (see Table 4.3). Therefore, according to our calculations, m-aminobenzoic acid molecules form thermodynamically stable cyclic carboxylic dimers in DMSO. These species correspond to the structural synthon found in the crystal structures II and V, the polymorphs where mABA exist in the non-zwiterionic form.

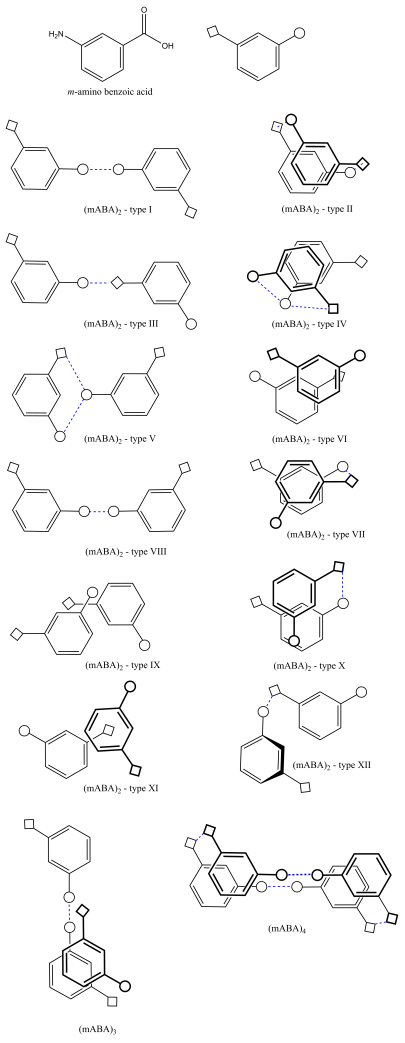

Figure 4.1: Schematic representation of m-aminobenzoic acid clusters, $(mABA)_n$ (n = 2-4), located in this study.

Figure 4.2 represents the coexistence profile of the free acid monomer, mABA, and dimer, $(mABA)_2$, in DMSO computed using the monomer-dimer model of Krishnan and co-workers to quantify the stoichiometric concentration of monomers and dimers in solution:[154]

$$[(mABA)_2] = \frac{\left\{ -1 + (1 + 8K_D[C]^{1/2}) \right\}^2}{16K_D} \tag{4.7}$$

In equation 4.7, K_D is the dimerisation constant, $K_D = [(mABA)_2]/[mABA]^2$, [C] is the stoichiometric concentration of the solute, [C] = [mABA] + 2[$(mABA)_2$], and [mABA] and [$(mABA)_2$] are the monomer and dimer concentrations, respectively (mol.L^{-1}). The value of K_D (304.9) has been computed from the Boltzmann average value of the free energies of dimerisation in DMSO (ΔG_{ass}^{*} = -3.40 kcal.mol^{-1}, Table 4.2). Fig. 4.2 shows that for [C] beyond 0.01 mol.L^{-1} (log[C] = 2) dimers of mABA, and in particular the classic carboxylic dimers I, are the predominant species in solution. However, this result is in contrast to what was reported by Hughes and co-workers, who, to explain changes in the ^{13}C NMR chemical shifts during the crystallisation process of mABA, suggested the formation in supersaturated solution of non-zwitterionic dimers linked by O-H...N hydrogen bonds. According to our static DFT calculations, none of the other H-bonded clusters is stable in DMSO. However, higher-order molecular clusters than the dimers could be stable in solution.

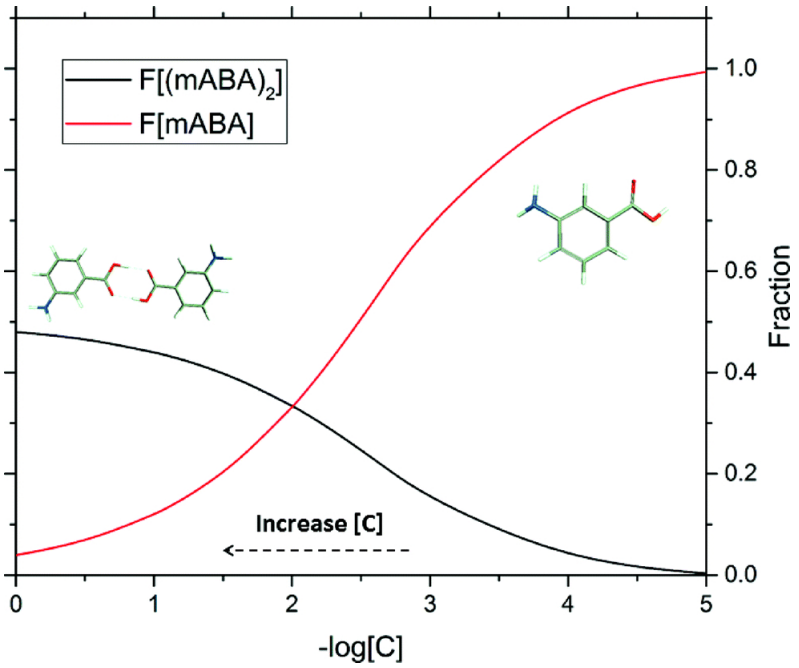

Figure 4.2: Fraction of the stoichiometric mABA concentration present as monomers and dimers in DMSO. Curve obtained by applying eqn 4.7 with $K_D = 304.9$ L.mol^{-1}.

4.3.2 Trimers

We modelled the formation of **trimers** $(mABA)_3$ in solution from the condensation reaction of the cyclic dimer I with a mABA molecule, $(mABA)_2$ + mABA $\rightarrow$ $(mABA)_3$, and the condensation of three isolated mABA molecules, 3 mABA $\rightarrow$ $(mABA)_3$. The geometries of the 10 low-lying $(mABA)_3$ structures were optimised and in solution, the Boltzmann averaged values of the free energy of the trimerization reaction 3 mABA $\rightarrow$ $(mABA)_3$ is positive (0.98 kcal.mol^{-1} in H_2O and 1.39 kcal.mol^{-1} in DMSO). However, these values are relatively small and we cannot exclude that traces of trimeric mABA species could be experimentally found in DMSO and account for the ^{13}C NMR chemical shifts observed during the crystallisation of mABA.[117]

4.3.3 Tetramers

For the tetramers, the structures and free energies of the most stable low-lying isomers of $(mABA)_4$ were determined from approximately 1500 candidate structures of type $(mABA)_4$, (D+D), which corresponds to tetramers generated from the aggregation of two cyclic dimers I. In fact, we verified that tetramers of type $(mABA)_4$, (D+D) are significantly lower in energy than tetramers of type $(mABA)_4$, (D+2M), which correspond to tetramers generated from the aggregation of dimer I with two isolated mABA molecules. A schematic representation of the most stable tetramer is reported in Fig 4.1 where $(mABA)_4$ consists of two dimeric units arranged in a stacked configuration; this promotes the stabilisation of the tetrameric structure by allowing the formation of both H-bonding and π-π interactions between pairs of meta-aminobenzoic acid molecules. The energetics of formation of $(mABA)_4$ (Table 4.2) indicate that in DMSO, where mABA molecules form stable classic carboxylic species, the aggregation of two dimers [$(mABA)_2$ + $(mABA)_2$ $\rightarrow$ $(mABA)_4$] is endergonic (+2.06 kcal.mol^{-1}). On the other hand, the formation of $(mABA)_4$ could result from the condensation of a single carboxylic dimer with two mABA monomers [$(mABA)_2$ + 2 mABA $\rightarrow$ $(mABA)_4$, -1.36 kcal.mol^{-1}] is negative. In water, the aggregation free energy of both tetramerisation pathways are negative (Table 4.2).

Therefore, starting from a large number of randomly generated candidate structures and by imposing a minimum free energy condition for the isomers of $(mABA)_4$, the most stable tetrameric species in solution were located. Fig. 4.3 highlights the correspondence between the most stable tetramer

and the $\pi - \pi$ stacking crystal synthon found in form II of mABA.[95] Consequently, the transfer of structural information from the solution- to the solid state-phase is not only related with the presence in solution of stable carboxylic acid dimers,[75,76,82] but the higher-order prenucleation cluster $(mABA)_4$ (Fig. 4.1) can also direct the nucleation of mABA towards the formation of the polymorphic form II.

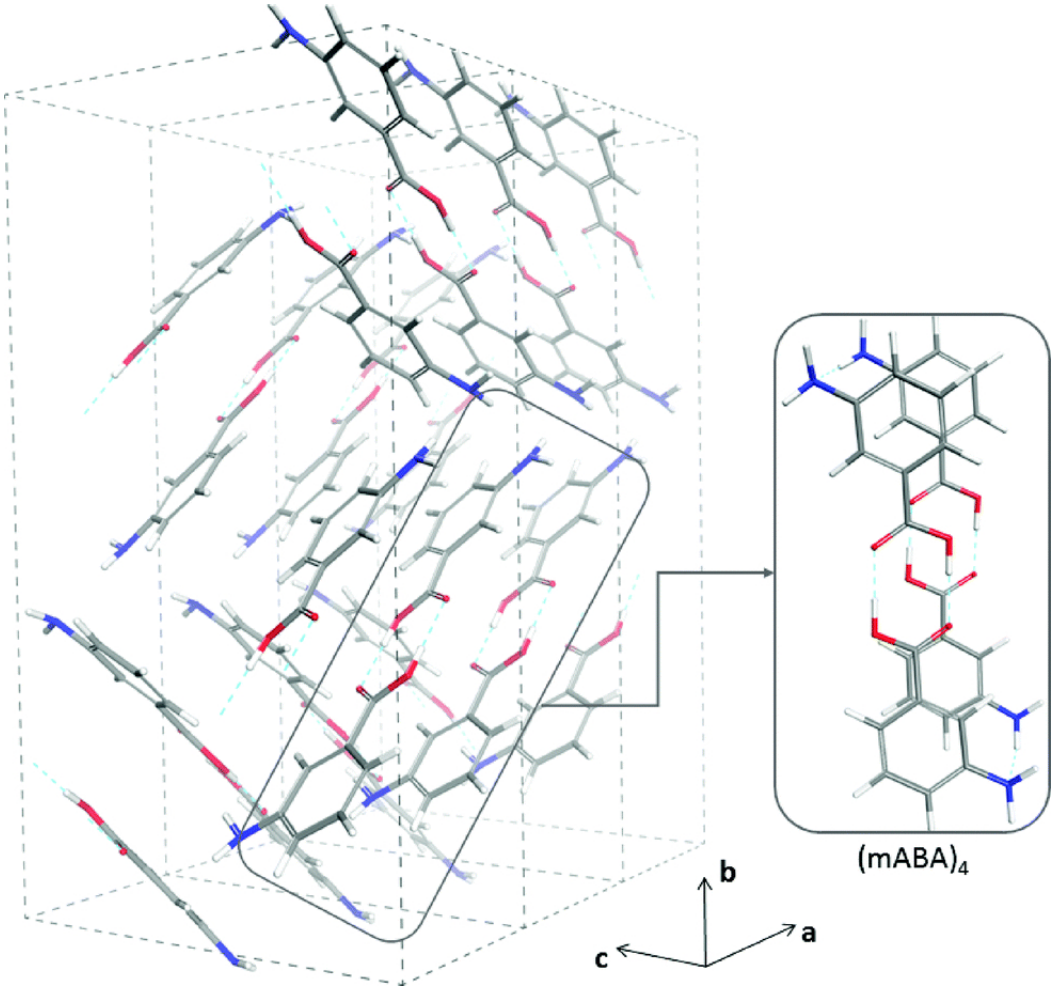

Figure 4.3: Correspondence between the most stable meta-aminobenzoic acid tetramer, $(mABA)_4$, in solution and the synthon found in the (1 x 3 x 1) unit cell of form II of mABA.

4.4 Role of solvent in the kinetics of aggregation

The quantum mechanical continuum calculations reported above describe the solution thermodynamics of the process of mABA cluster formation with respect to infinitely separated mABA molecules, but ignore intermediate processes such as separation of solute-solvent molecules and diffusion of solute molecules across the solution.

Figure 4.4 displays the last configurations of the MD simulations of mABA solutions in water (20 ns) and DMSO (200 ns). The results indicate a clear solvent-dependent behaviour for the aggregation of mABA. In water, even at low concentrations (0.3 mol.L^{-1}), π-π associated clusters form spontaneously, whereas in DMSO the mABA molecules are present in a more solvated state; the level of clustering in water is therefore significantly larger than in DMSO. For the DMSO solutions, extending the simulation period to values larger than 200 ns did not results in a significant increase of the level of aggregation or to the formation of the thermodynamically stable cyclic dimer I predicted by the DFT solvation continuum calculations (see Table 4.2).

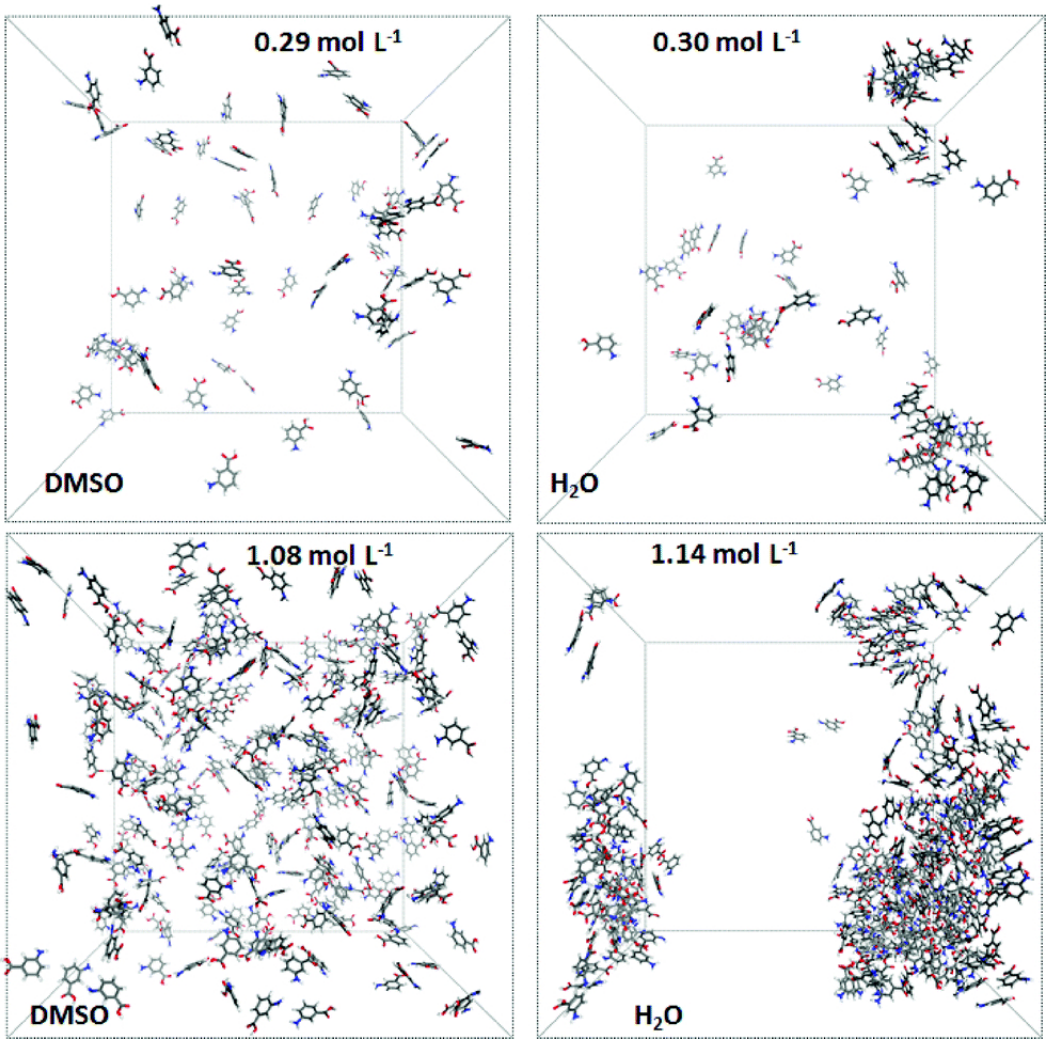

Figure 4.4: Last configurations of MD simulations of 0.3 mol.L^{-1} and 1.1 mol.L^{-1}. mABA solutions in DMSO (200 ns) and water (30 ns). For clarity, solvent molecules were removed.

We quantified the level of aggregation of mABA in water and DMSO by computing, every 10 ps, the number of mABA pairs within 3.5 Å (Fig. 4.5) and the number of hydrogen bonds between mABA molecules. As expected, the number of hydrogen-bonded clusters increases with the concentration of the solute, but the role of solvent in assisting, or hindering, the process of cluster formation is substantial. For example, in the 0.3 mol.L^{-1} and 1.1 mol.L^{-1} aqueous solutions the average number of pairs in water are

approximately 50% and 130% higher than in DMSO, respectively (see Fig. 4.5). Notice also the small temporal fluctuations for the number of mABA pairs in DMSO (after 200 ns) with its equivalent concentration in water (after 10 ns), which indicates that the clustering of mABA in water is still progressing and is likely to be considerably higher than the one reported in figure 4.5.

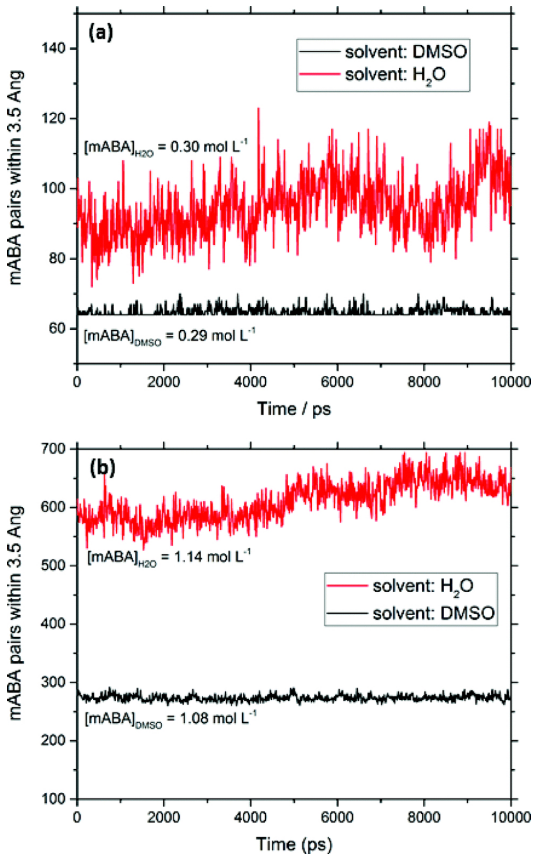

Figure 4.5: Time evolution of the number of pairs between mABA molecules computed during the last 10 ns of the MD simulations of (a) 0.3 mol.L^{-1} and (b) 1.1 mol.L^{-1} solutions.

4.5 Free energy profiles of dimerisation

Figure 4.6 shows free energy profiles for the dimerisation of mABA in DMSO and water. In DMSO, the carboxylic dimer I (A in Fig. 4.6) is stable with respect to highly separated (40 Å) mABA molecules and thermodynamically more stable than in water, which is in agreement with the DFT values for free energies of dimerisation (see Table 4.2). As these two units are brought together during the metadynamics simulations, the free energy profiles in DMSO and water have significant differences even at large distances between the mABA molecules indicating a different desolvation behaviour of mABA in the two solvents. In particular, the formation of $(mABA)_2$ goes through two highly solvated intermediate states, D and C, and an activated state B.

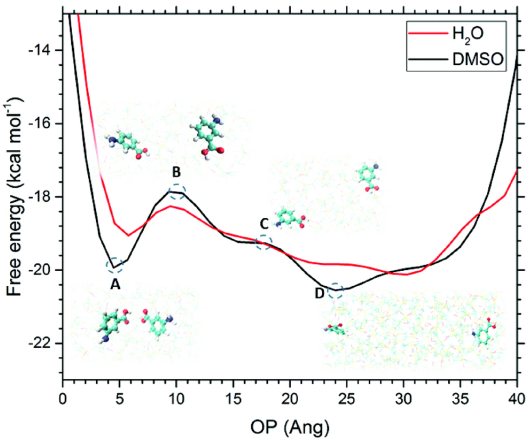

Figure 4.6: Free energy profiles for the meta-aminobenzoic acid dimerisation in water and DMSO. The order parameter (OP) used to study this reaction was defined by averaging two distances between the hydroxyl hydrogen and carboxyl oxygen of a pair of mABA molecules. A, B, C and D represent the structure of the microstates along the OP trajectory determined using the *METAGUI* computational tool.[155]

In aqueous solution, the formation of $(mABA)_2$ is less thermodynamically favourable than in DMSO, in agreement with the quantum mechanical continuum calculations, but the free energy profile of mABA dimerisation has a significantly lower energy barrier. On the other hand, in DMSO the activation free energy required for the formation of the mABA classic carboxylic dimer (3 kcal.mol^{-1}) is approximately twice the activation free energy in water and significantly larger than kT (0.593 kcal.mol^{-1} at 300 K), which rationalises the absence of classic carboxylic acid dimers during unbiased MD simulations in DMSO solutions (see Fig. 4.4).

We have also computed the free energy profile with respect to the carbon (C=O) carbon (C=O) distance of two mABA molecules (Fig. 4.7).

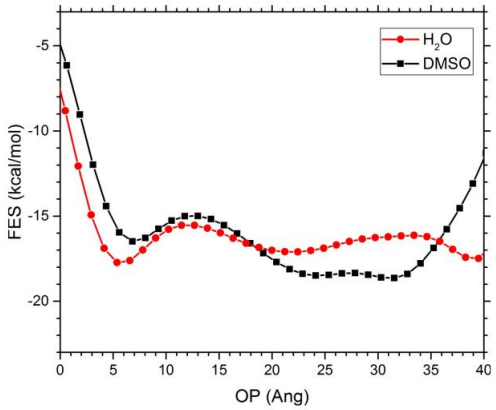

Figure 4.7: Free energy profile for the mABA dimerisation reaction in water and DMSO obtained by using metadynamics. The order parameter (OP) used to study the dimerisation reaction was the carbon (C=O) carbon (C=O) distance of two mABA molecules.

This order parameter leads to the formation of an H-bonded π-π dimer. In DMSO, the activation free energy is approximately 3.5 kcal.mol^{-1} whereas in water the formation of a π-π dimer is almost barrier-less, again in agreement with the spontaneous formation of π-π clusters during unbiased MD simulations.

These results confirm that explicit solute-solvent interactions, which are critical during the process of mABA desolvation (Chapter 3), control also the kinetics of molecular dimerisation from solution and, consequently, the formation of larger pre-nucleation clusters.

This result also supports what was previously suggested by Rasmuson and co-workers that the strength of the solute-solvent molecular interaction is directly related with experimental rate of nucleation of organic molecules from solution.

4.6 Conclusions

Computer simulations based on density functional theory, molecular dynamics and metadynamics methods have revealed the role of solvent in the kinetics and thermodynamics of aggregation of mABA molecules in aqueous and organosulfur (DMSO) solutions.

DFT calculations with a continuum model to describe the solvation environment were used to probe the potential energy surfaces of molecular clusters of meta-aminobenzoic acid, $(\text{mABA})_n$ (n = 2-4), locate the low-lying energy structures of $(\text{mABA})_n$, and compute their Gibbs free energies in water and DMSO. Starting from many randomly generated candidate structures and by imposing the condition of minimum free energy in solution for the isomers of $(\text{mABA})_n$, we proved that the structure of the most stable dimer and tetramer correspond to the classic carboxylic dimer π-π stacking synthon found in the crystalline form II of mABA. Consequently, the transfer of structural information from the solution- to the solid state-phase of mABA is not only related with the presence of stable carboxylic

acid dimers, as previously suggested, but also to higher-order prenucleation clusters $(mABA)_4$, which could also direct the nucleation process towards the formation of the polymorphic form II of mABA.

MD simulations of mABA solutions show a significant solvent-dependent behavior for the aggregation of mABA molecules. In aqueous solutions, even at relatively low concentrations, mABA molecules spontaneously form H-bonded π-π stacking cluster. In organosulfur solutions, molecules of mABA are in a more solvated state and high concentrations of mABA are required to observe appreciable levels of mABA aggregation, but the formation of the stable dimeric and tetrameric species predicted by the quantum mechanical continuum solvation calculations could not be observed during the MD simulation.

Free energy profiles for the mABA dimerisation computed using metadynamics simulations show that in DMSO there is a higher Gibbs activation energy associated with the diffusion and desolvation of mABA, which are necessary and pre-requisite steps for the aggregation of organic molecules from solution to occur. In particular, the formation of $(mABA)_2$ in DMSO has an activation barrier that is twice the one in water, and much larger than the value of kT at 300 K. This rationalises the observed solvent-dependent aggregation behavior of mABA and agrees with the DFT calculations of the formation of microsolvated mABA clusters, $mABA(S)_n$, reported in Chapter 3, which show that mABA-DMSO interaction is substantially larger (up to 10 kcal.mol^{-1}) than mABA-water interaction.

Chapter 5

Aggregation of Meta-aminobenzoic Acid in Water

Meta-aminobenzoic acid exists in water in both the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) forms. However, the constituent molecules of the polymorph that crystallises from aqueous solutions are zwitterionic. This chapter reports atomistic simulations of the events surrounding the early stage of crystal nucleation of meta-aminobenzoic acid from aqueous solutions. Density functional theory calculations were conducted to determine the low-lying energy conformers of meta-aminobenzoic acid dimers and to compute the Gibbs free energies in water of nonionic, $(mABA)_2$, zwitterionic, $(mABA^{\pm})_2$ and nonionic-zwitterionic, $(mABA)(mABA^{\pm})$, species. Classical molecular dynamics simulations of mixed mABA-$mABA^{\pm}$ aqueous solutions were carried out to examine the aggregation of meta-aminobenzoic acid. According to these simulations, the selective crystallisation of the

polymorph, with constituent molecules in the zwitterionic form, is driven by the formation of zwitterionic dimers in solution, which are thermodynamically more stable than $(mABA)_2$ and $(mABA)(mABA^{\pm})$ pairs. This work represents a paradigm of the role of molecular processes during the early stages of crystal nucleation in affecting polymorph selection during crystallisation from solution.

5.1 Introduction

The very strong polymorphic character of meta-aminobenzoic acid is related to the manifold of inter-molecular interactions between meta-aminobenzoic acid molecules (hydrogen (H) bonding, π-π interactions and H-π interactions) but also to the ability of this molecule to exist in either of both the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) forms.[113] The constituent molecules of the polymorphs denoted I, III and IV are zwitterionic, whereas in the polymorphs II and V they are nonionic:[95,99] in II, mABA molecules interact through the O-H...O acid dimer of a $R_2^2(8)$ ring motif [Figure 5.1(a)]; in III, $mABA^{\pm}$ molecules form ionic N^+-H...O^- interactions in a $R_4^4(8)$ ring motif [Figure 5.1(b)]; in Form IV, two independent molecules form a linear C(7) chain through ionic N^+-H...O^- interactions [(Figure 5.1(c)]. The crystal structure of Form I has not been determined so far and the crystal structure of Form V shows disorder.[95,99,156]

The nature of the solvent can significantly control the formation of one specific polymorph over another.[114,157] Form II preferentially crystallises from dimethyl sulfoxide (DMSO),[99] where meta-aminobenzoic acid exists in the nonionic form.[117,157] On the other hand, Form I preferentially crys-

tallises from aqueous environments,[99] despite the values of the equilibrium constant $K_Z = [mABA^{\pm}]/[mABA]$ for aminobenzoic acids are of the order of unity in water,[97,98,158] implying a comparable distribution of $mABA^{\pm}$ and mABA molecules. The fundamental details of factors controlling the selection between zwitterionic and nonionic forms of meta aminobenzoic acid during crystal nucleation from aqueous solution are not known yet.[117] This work aims to solve this conundrum by applying a combination of atomistic methods to follow the events surrounding the crystal nucleation of meta-aminobenzoic acid from aqueous solutions. Density functional theory (DFT) calculations have been used to determine the structure and energetics of formation in water of $(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$ dimers. Classical molecular dynamics (MD) simulations of mixed mABA-$mABA^{\pm}$ aqueous solutions have been conducted to quantify the aggregation of meta-aminobenzoic acid.

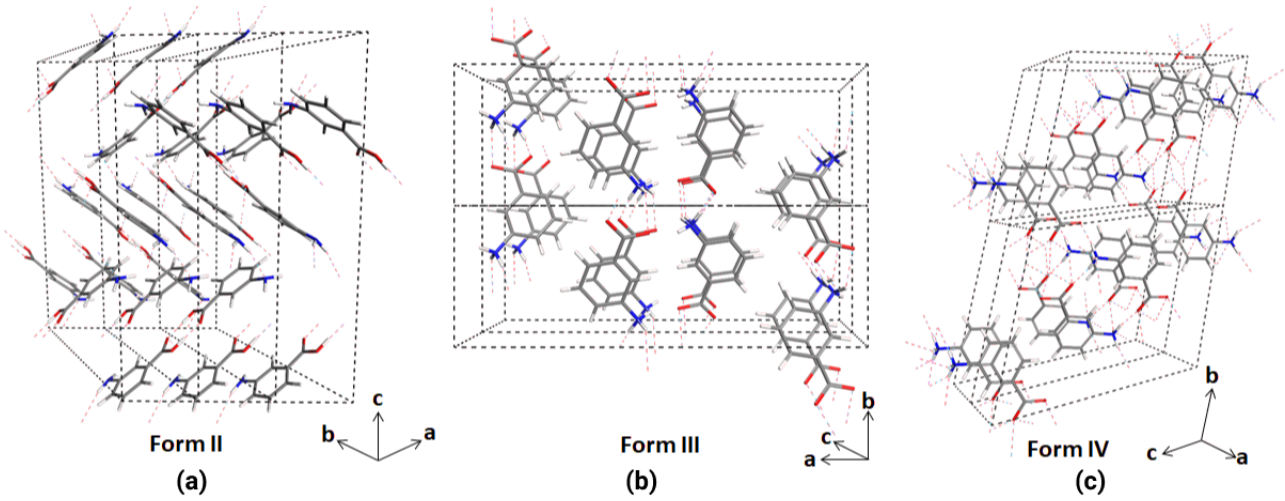

Figure 5.1: Crystal structure of the polymorphs of meta-aminobenzoic acid denoted II, III and IV: (a) (1x3x1) unit cell of Form II (neutral); (b) (1x2x2) unit cell of Form III (zwitterionic); (c) (2x2x1) unit cell of Form IV (zwitterionic).[95]

5.2 Computational methods

5.2.1 Density functional theory calculations

Electronic structure calculations were carried out with the NWChem (version 6.3) and Gaussian09 codes. The Grimme's density functional (B97-D) and the Minnesota 06 global hybrid functional with 54% Hartree-Fock (HF) exchange (M06-2X) were used together with the Gaussian 6-31+G(d,p) basis set. Free energies of solvation were calculated using the SMD solvation model. The free energies of formation of nonionic, $(mABA)_2$, nonionic-zwitterionic, $(mABA)(mABA^{\pm})$, and zwitterionic, $(mABA^{\pm})_2$, dimers were computed according to the following equation:

$$\Delta G^*_{ass} = G^*_{AB} - G^*_A - G^*_B \tag{5.1}$$

In Eq. 5.1, G^*_X is the total Gibbs free energy of the species X (X = AB, A or B) in the liquid. This quantity was evaluated using two different approaches. The first one follows the recommendation by Ho *et al.* that free energies of molecules in solution should be obtained from separate gas- and solution-phase calculations;[159] the following expression was used to evaluate the Gibbs free energy of the species X:

$$G^*_X = E_{e,gas} + \delta G^0_{VRT,gas} + \Delta G^*_{solv} RTln[\tilde{R}T] \tag{5.2}$$

In Eq. 5.2, $E_{e,gas}$ is the gas-phase total electronic energy of the gas-phase optimised geometry of the species X, $\delta G^0_{VRT,gas}$ is the vibrational-rotational-translational contribution to the gas-phase Gibbs free energy at T = 298 K under a standard-state partial pressure of 1 atm, ΔG^*_{solv} is the

solvation free energy of the solute corresponding to transfer from an ideal gas at a concentration of 1 mol.L^{-1} to an ideal solution at a liquid-phase concentration of 1 mol.L^{-1}, and the last term is the free energy change of 1 mol of an ideal gas from 1 atm to 1 mol.L^{-1} (RTln[$\tilde{R}$T] = 1.89 kcal.mol^{-1} at 298 K, $\tilde{R}$ = 0.082 K^{-1}).[108] However, the gas-phase optimisation of zwitterionic, (mABA$^{\pm}$)$_2$, and nonionic-zwitterionic, (mABA)(mABA$^{\pm}$), dimers caused the H-transfer between molecular units (*e.g.* (mABA$^{\pm}$)$_2$ $\rightarrow$ (mABA)$_2$). In these instances, stationary points in the solution do not correspond to stationary points in gas-phase, making it impossible to compute relevant gas-phase vibrational, translational and rotational contributions ($\delta G^{0}_{VRT,gas}$). The other approach adopted was to optimise the structures of (mABA$^{\pm}$)$_2$, (mABA)(mABA$^{\pm}$), and of the monomers mABA and mABA$^{\pm}$, in the aqueous phase; the following expression was then used to evaluate the free energy of the species:

$$G^{*}_{X} = E^{Tot}_{soln} + \delta G^{*}_{VRT,soln} \tag{5.3}$$

where $\delta G^{*}_{VRT,soln}$ is the vibrational-rotational-translational contribution to the liquid-phase, and E^{Tot}_{soln} is given by the sum of the liquid-phase expectation value of the gas-phase Hamiltonian ($E_{e,soln}$), the electronic polarisation contribution to the solvation free energy based on bulk electrostatic (ΔG_{EP}), and the contribution from cavitation, dispersion and solvent structural effects (G_{CDS}):

$$E^{Tot}_{soln} = E_{e,soln} + \Delta G_{EP} + G_{CDS} \tag{5.4}$$

The potential energy surface of a molecular cluster is characterised

by multiple low-lying energy isomers.[160] The free energy of the dimers $(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$ was therefore determined from the Boltzmann ensemble average:

$$\langle G(X) \rangle = \sum_{i=1}^{N} f_i G(X_i) \tag{5.5}$$

where f_i is the Boltzmann factor corresponding to the i^{th} configuration, $G(X_i)$ is the corresponding free energy and N is the number of low-lying energy isomers. The Boltzmann factor was determined according to:

$$f_i = \frac{e^{-G(X_i)/RT}}{\sum_j e^{-G(X_j)/RT}} \tag{5.6}$$

where R is the universal gas constant, T is the absolute temperature ($T = 298$ K) and the index j runs over all isomers. The low-lying energy structures of the meta-aminobenzoic acid dimers were located using the following computational protocol: (1) For each type of dimer [$(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$] hundreds of thousands of candidate structures were generated using Granada,[111,161] a code designed to distribute randomly one or more molecules around a central unit (a monomer, dimer, trimer etc.) placed at the center of a cube of defined side length. (2) Configurations satisfying the condition that at least one atom of each mobile molecule was within 4 Å from at least one atom of the central unit were selected as potential low-lying energy structures. (3) The energies of these structures were evaluated at the B97-D/6-31+G(d,p) level of theory and the Boltzmann factor f_i corresponding to the i^{th} configuration was determined as:

$$f_i = \frac{e^{-(E_i - E_0)/RT}}{\sum_j e^{-(E_j - E_0)/RT}} \tag{5.7}$$

where E_i was the energy of the i^{th} candidate structure and E_0 was the energy of the most stable candidate structure. (4) The candidate structures with a Boltzmann factor $f_i \geq 0.01$ and ten to fifteen randomly selected structure such that $3 \leq E_i - E_0 \leq 15$ kcal.mol^{-1} were selected. (4) Geometry optimisation, thermochemical properties and solvation energies of the selected configurations were computed at the M06-2X/6-31+G(d,p) level of theory.

5.2.2 Molecular dynamics simulations

Classical MD simulations were performed using version 5.0.4 of the GROMACS molecular dynamics package. The leapfrog algorithm with a time step of 2 fs was used to integrate the equations of motion. The isothermal-isobaric (constant NPT) ensemble was used to maintain a temperature of 300 K and a pressure of 1 bar. The velocity rescale thermostat and the isotropic Parrinello-Rahman barostat were used with 0.4 ps and 2.0 ps as the thermostat and barostat relaxation times, respectively. The electrostatic forces were calculated by means of the particle-mesh Edwald approach with a cutoff of 1.2 nm. The same cutoff was used for the van der Waals forces. The LINCS algorithm was applied at each step to preserve the bond lengths. The general AMBER forcefield (GAFF) was used to model the nonionic and zwitterionic (mABA$^{\pm}$) forms of meta-aminobenzoic acid; this family of forcefields has been previously used to compute the aggregation and crystal growth of organic molecules.[92,157,162,163] Water molecules were modelled using the SPC/E potential.[142] The interactions between

mABA and $mABA^{\pm}$ molecules and between these molecules and water were described using the GAFF potential. To generate the GAFF parameters for mABA and $mABA^{\pm}$, the structure and molecular electrostatic potential of these molecules were computed using the HF method and the 6-31G* basis set, and the Antechamber package was then used to compute partial charges according to the restrained electrostatic potential formalism. The GAFF forcefields and partial charges of mABA and $mABA^{\pm}$ are reported in the Appendix A.2.

The *insert-molecules* utility of GROMACS was used to generate aqueous meta-aminobenzoic acid solutions of different concentrations by inserting equal amounts of mABA and $mABA^{\pm}$ molecules in an empty cubic box of size 5 nm. The *solvate* utility was then used to solvate the cubic boxes with SPC/E water. Each solution was at first minimised using the conjugate-gradient algorithm with a tolerance on the maximum force of 200 $kJ.mol^{-1}$, and the temperature and volume of each system were equilibrated by running 100 ps of constant volume, constant temperature (NVT) simulation followed by 200 ns of NPT simulations. Analysis were conducted on the last 40 ns of simulation. Details of the simulation times, number of solute and solvent molecules, equilibrated values of the average cell length are reported in Table 5.1. The convergence of the box cell volume during the period of equilibration are shown in Figure 5.2.

Table 5.1: Details of molecular dynamics simulation.

no. of mABA molecules	no of $mABA^{\pm}$ molecules	no. of H_20 molecules	Box length (Å)	Conc. (mol/L)	Simulation time (ps)
4	4	11385	69.89	0.04	400000
8	8	11230	69.69	0.08	200000
16	16	11123	69.75	0.16	200000
32	32	11101	69.95	0.31	200000
64	64	11101	70.67	0.60	200000

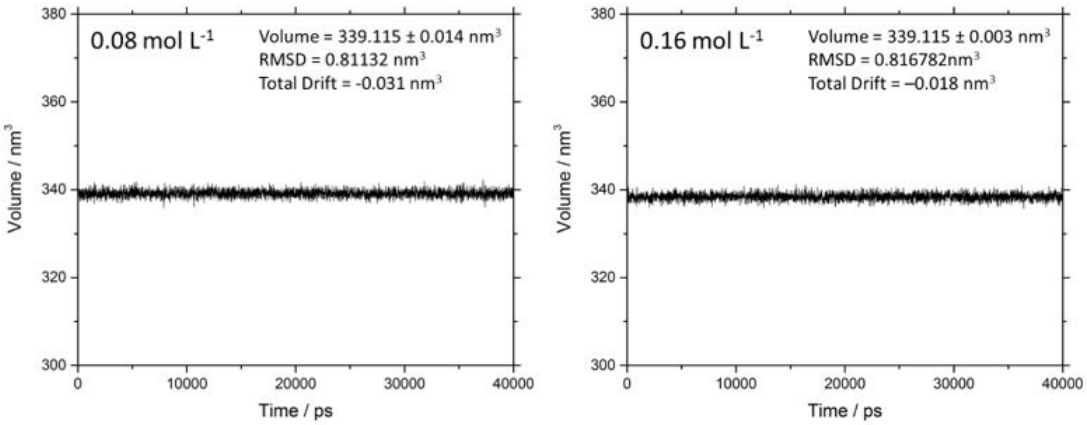

Figure 5.2: Convergence of the volume of the simulation boxes containing mixed mABA-$mABA^{\pm}$ solutions.

5.3 Dimerisation of meta-aminobenzoic acid

Stable dimers in solution have often been linked to the structural synthon found in the polymorph that crystallises from solution.[74,164] This section reports therefore results from extensive DFT calculations to determine the structure and the thermodynamic stability in water of meta-aminobenzoic acid dimers. The Boltzmann averaged energetics of formation of the nonionic, $(mABA)_2$, zwitterionic, $(mABA^{\pm})_2$, and nonionic-zwitterionic, $[(mABA)(mABA^{\pm})]$, dimers are reported in Table 5.2. The free energy of formation of $(mABA)_2$ ranges from -0.1 to 2.4 kcal.mol^{-1},

depending on the method used to compute the total free energies of the dimers and monomer in water. The formation of (mABA)(mABA$^{\pm}$) (2.4 kcal.mol^{-1}) is also endergonic. On the other hand, the dimerisation free energy of the zwitterionic dimer (mABA$^{\pm}$)$_2$ is large and negative (-5.8 kcal.mol^{-1}).

Table 5.2: Energetics of dimerisation of meta-aminobenzoic acid: $\Delta E_{e,gas}$ is the gas phase interaction energy; ΔG^0_{ass} is the standard state (1 atm) gas-phase association free energy at 298 K; ΔG^*_{ass} is the standard state (1 mol.L^{-1}) free energy of reaction in the liquid-phase. Calculations conducted at the M06-2X/6-31+G(d,p) level of theory using the SMD solvation model. Values obtained from the Boltzmann average of the energies, or free energies, of the isomers of nonionic (mABA)$_2$, zwitterionic (mABA$^{\pm}$)$_2$, and mixed (mABA)(mABA$^{\pm}$) dimers. Values in kcal.mol^{-1}.

Reaction	Δ E$_{e,gas}$	Δ G$^0_{ass}$	Δ G$^*_{ass}$
2 mABA $\rightarrow$ (mABA)$_2$	-18.3	-6.6	-0.1[1]
			2.4[2]
mABA + mABA$^{\pm}$ $\rightarrow$ (mABA)(mABA$^{\pm}$)			1.3[2]
2 mABA$^{\pm}$ $\rightarrow$ (mABA$^{\pm}$)$_2$	-	-	-5.8[2]

[1]Gas-phase optimised geometries and free energies in water obtained using Equation 5.2;

[2]Solution-phase optimised geometries and free energies in water obtained using Equation 5.3.

Figure 5.3 reports the structures of the thermodynamically most stable (mABA)$_2$ and (mABA)(mABA$^{\pm}$) species in water. The (mABA)$_2$ dimer corresponds to the structural synthon found in Form II,[95] where the two nonionic meta-aminobenzoic acid molecules interact through a double H-bond to form a classic carboxylic dimer (Figure 5.3(a)). In the (mABA)(mABA$^{\pm}$) dimer, the two monomers are arranged to maximise the concomitant H-bonding and π-π interactions (Figure 5.3(b)). All other (mABA)$_2$ and (mABA)(mABA$^{\pm}$) dimeric structure have significantly

higher free energies of formation in water (2.5 kcal.mol^{-1} $< \Delta G^*_{ass} <$ 10 kcal.mol^{-1}) and they are, therefore, unstable in aqueous solution. On the other hand, several stable zwitterionic dimers, $(mABA^{\pm})_2$, were found in solution (Figure 5.3(c)). Consequently, even though the distribution between zwitterions and nonionic molecules in water is close to unity,[97?,98] we make the hypothesis that the selective crystallisation of the polymorphs that only contain zwitterionic molecules (Forms I, III, and V) is driven by the higher thermodynamic stability in water of zwitterionic $(mABA^{\pm})_2$ dimers.

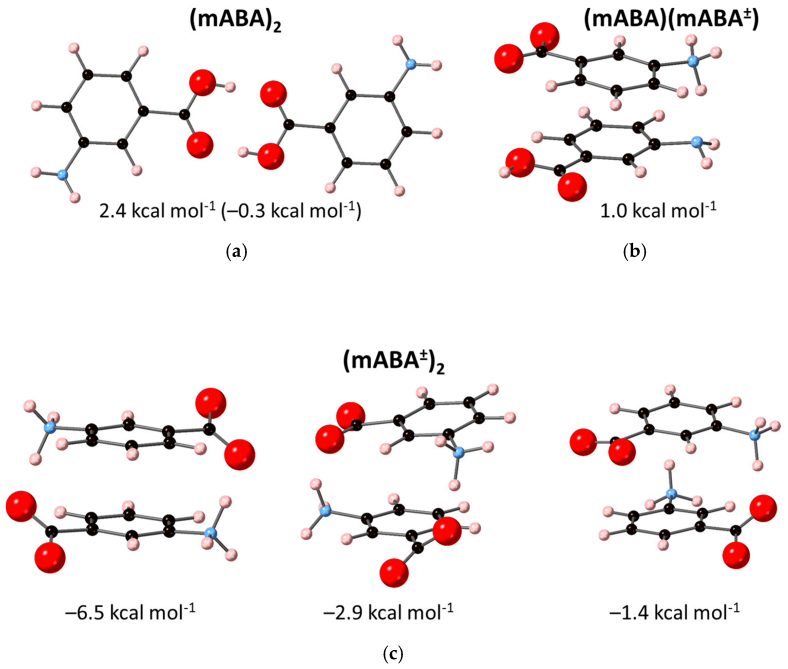

Figure 5.3: Optimised structures of most stable meta-aminobenzoic acid dimers in water: (a) nonionic $(mABA)_2$ dimer (in parenthesis value obtained using the gas-phase optimised geometries of $(mABA)_2$ and mABA); (b) nonionic-zwitterionic $(mABA)(mABA^{\pm})$ dimer; (c) zwitterionic $(mABA^{\pm})_2$ dimer. Beneath the structure is reported free energy of dimer formation in water.

5.4 Molecular aggregation in mixed mABA/mABA$^{\pm}$ aqueous solutions

Classical MD simulations ($\geq$ 200 ns) of mixed mABA-mABA$^{\pm}$ aqueous solutions were conducted to examine the aggregation behavior of meta-aminobenzoic acid as a function of concentration. Four concentrations were considered: 0.04 mol.L^{-1}, 0.08 mol.L^{-1}, 0.16 mol.L^{-1}, and 0.31 mol.L^{-1}. Svard *et al.*[99] reported crystallisation experiments of meta-aminobenzoic acid at saturated solution. At 300 K, the solubility in water of polymorph I is 5.4 g.L^{-1},[94] whereas for the other polymorphs higher solubility values have been reported: 7.8 g.L^{-1} for Form II, 6.07 g.L^{-1} for Form III, and 6.25 g.L^{-1} for Form III.[156] The 0.04 mol.L^{-1} solution (5.3 g.L^{-1}) corresponds therefore to conditions just below the solubility limit of Form I, while the others simulated systems (10.8 g.L^{-1}, 21.5 g.L^{-1}, and 42.6 g.L^{-1}) correspond to increasingly supersaturated solutions with respect to all polymorphs of meta-aminobenzoic acid. Representative configurations of these solutions are reported in Figure 5.4, where the number of molecular aggregates that form in solution increases as a function of solute concentration. This aggregation process has been quantified in terms of the number of (mABA...mABA), (mABA$^{\pm}$...mABA$^{\pm}$), and (mABA...mABA$^{\pm}$) pairs within 4.0 Å (Figure 5.5 and Figure 5.6). The number of molecular pairs increases with the concentration but the number of nonionic clusters is significantly higher than mixed and zwitterionic species. As the dehydration of the molecules of solute is a crucial step during crystal nucleation from solution,[79] the stronger interaction of mABA$^{\pm}$ with the surrounding water molecules discussed in Chapter 3 could explain the observed different level

of aggregation of nonionic and zwitterionic species in water. Moreover, a close view of the clusters formed during the MD simulations reveals that meta-aminobenzoic acid interact via a manifold of inter-molecular interactions: H-bonding X-H...X (X = O or N) between amino (NH_2 and NH_3^+) and carboxylic (COOH and COO^-) groups, π-π interactions between benzine (C_6H_4) groups, and X-H...π interactions.

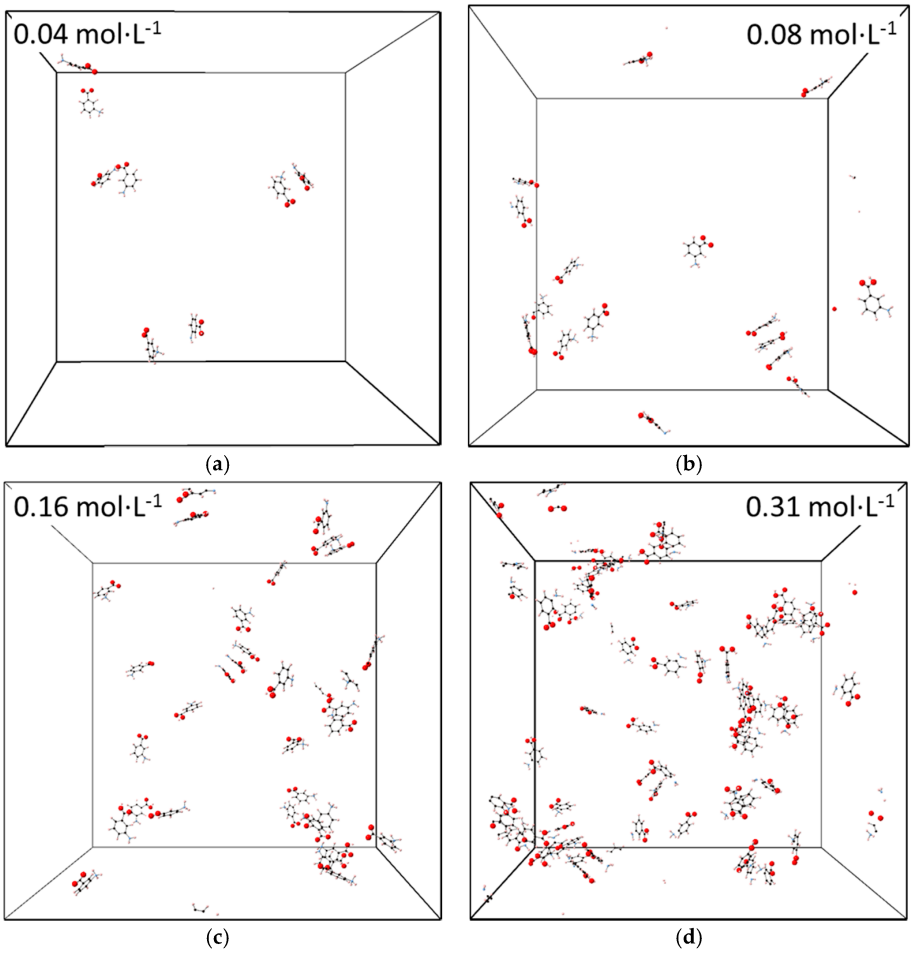

Figure 5.4: Configuration at 200 ns of mixed mABA-mABA$^{\pm}$ aqueous solutions: (a) 0.04 mol.L^{-1} aqueous solution; (b) 0.09 mol.L^{-1} aqueous solution; (c) 0.16 mol.L^{-1} aqueous solution; (d) 0.31 mol.L^{-1} aqueous solution. Water molecules have been removed. The grey outlines represent the cubic simulation box.

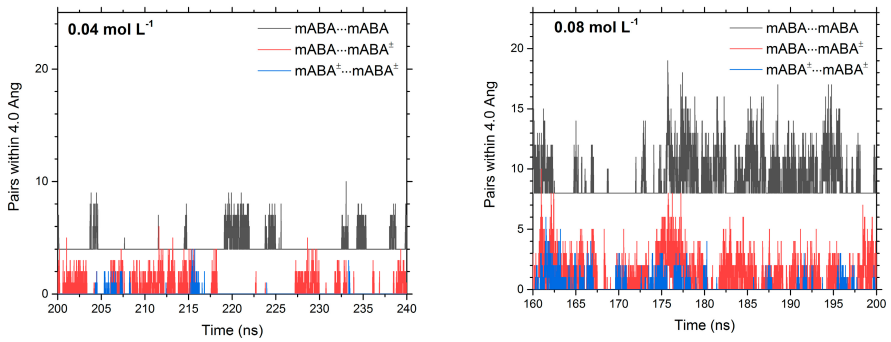

Figure 5.5: Time evolution of the number of pairs between meta-aminobenzoic acid molecules in mixed mABA-mABA$^{\pm}$ aqueous solutions computed during the last 40 ns of the MD simulations.

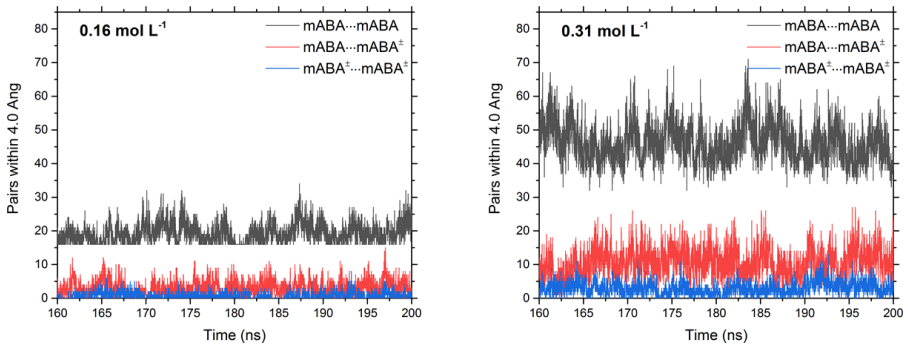

Figure 5.6: Time evolution of the number of pairs between meta-aminobenzoic acid molecules in mixed mABA-mABA$^{\pm}$ aqueous solutions computed during the last 40 ns of the MD simulations: (a) 0.16 mol.L^{-1}; (b) 0.31 mol.L^{-1}.

To characterise these interactions, a three-body simplified representation of the nonionic mABA (A-B-C) and zwitterionic mABA$^{\pm}$ (A*-B*-C*) molecules have been adopted (Figure 5.7), where A and A* represent the center-of-masses of -NH_2 and -NH_3^+, B and B* represent the center-of-masses of the benzine (C_6H_4) groups, and C and C* represent the center-of-masses of -COOH and -COO^-.

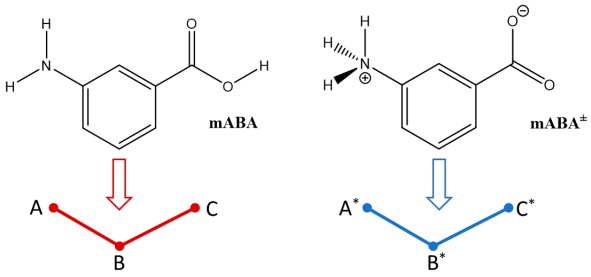

Figure 5.7: Three-body representations (A-B-C) and (A*-B*-C*) of the nonionic, mABA, and zwitterionic, mABA$^{\pm}$ forms of meta-aminobenzoic acid: A and A* are the center-of-masses (COMs) of the -NH_2 and -NH_3^+ groups, B and B* are the COMs of the benzine (C_6H_4) group, and C and C* are the COMs of the -COOH and -COO^- groups.

A symmetric pairwise interaction matrix (PIM) can therefore be used to quantify the interactions between (A-B-C) and (A*-B*-C*):

$$PIM = \begin{bmatrix} p_{A^*A^*} & p_{A^*B^*} & p_{A^*C^*} & p_{A^*A} & p_{A^*B} & p_{A^*C} \\ & p_{A^*B} & p_{B^*C^*} & p_{B^*A} & p_{B^*B} & p_{B^*C} \\ & & p_{C^*C^*} & p_{C^*A} & p_{C^*B} & p_{C^*C} \\ & & & p_{AA} & p_{AB} & p_{AC} \\ & & & & p_{BB} & p_{BC} \\ & & & & & p_{CC} \end{bmatrix}$$

The elements of the PIM matrix are defined as

$$p_{ij} = \langle \sum_i \sum_{i>j} f(r_{ij}) \rangle \tag{5.8}$$

where the pairwise interaction function f(r_{ij}) quantifies the existence of a (i,j) pair within a cutoff distance of 4.0 Å:

$$f(r_{ij}) = \begin{cases} 0, & r_{ij} > 4.0\text{Å} \\ 1, & r_{ij} < 4.0\text{Å} \end{cases} \tag{5.9}$$

The cutoff value of 4.0 Å was based on the analysis of the intermolecular distances between amino (NH_2 and NH_3^+), carboxylic (COOH and COO^-), and benzine (C_6H_4) groups in the most thermodynamically stable $(mABA)_2$, $(mABA)(mABA^{\pm})$, and $(mABA^{\pm})_2$ dimers in water (Figure 5.8).

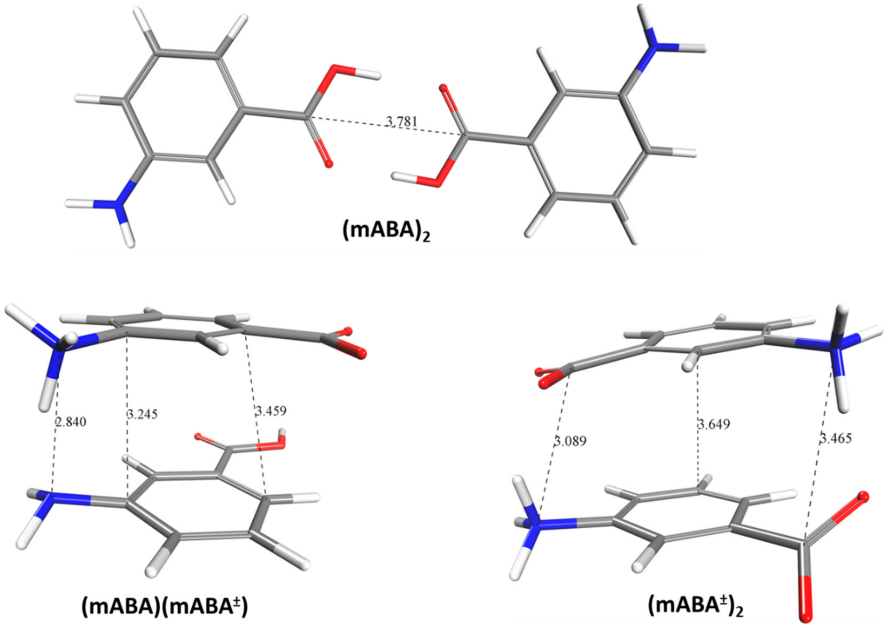

Figure 5.8: Intermolecular distances between the amino (NH_2 and NH_3^+), carboxylic (COOH and COO^-), and benzine (C_6H_4) groups in the most thermodynamically stable $(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$ dimers in water.

For example, the element p_{AA} corresponds to COOH...COOH interactions found in the classic carboxylic dimer $(mABA)_2$ (Figure 5.3(a)), the elements p_{A^*A} and p_{C^*C} correspond to the COO^-...COOH and NH_3^+...NH_2 interactions in the nonionic-zwitterionic dimer $(mABA)(mABA^{\pm})$ (Figure 5.3(b)), and the elements $p_{B^*B^*}$ and $p_{A^*C^*}$ correspond to $\pi...\pi$ and NH_3^+...COO^- interacting pairs in the structures of the most stable zwitterionic dimers $(mABA^{\pm})_2$ (Figure 5.3(c)). For the mixed 0.08 mol.L^{-1}

mABA-mABA$^{\pm}$ aqueous solutions, the pairwise interaction matrix in Table 4 reveals a higher proportion of NH_3^+...COO^- (A*...C* = 8.7%) and π...π (B*...B* = 9.1%) pairs than COOH...COOH (C...C = 6.5%), COO^-...COOH (C*...C* = 6.5%) and NH_3^+...NH_2 (A*...A = 5.3%). Very similar PIM matrices were obtained from the calculation of the three-body pairwise interactions of the other systems (Tables A.1, A.2, A.3 in the Appendix). This analysis implies that aqueous solutions of meta-aminobenzoic acid contain a higher proportion of stable zwitterionic $(mABA^{\pm})_2$ pairs, in agreement with the DFT calculations of dimerisation free energies.

Table 5.3: Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.08 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage.

	A*	B*	C*	A	B	C
A*	0.2	0.6	8.7	5.3	3.4	3.5
B*		9.1	2.7	2.6	10.0	7.6
C*			0.1	3.6	2.3	3.6
A				4.3	4.2	5.2
B					6.1	10.6
C						6.5

5.5 Conclusions

The aggregation of the nonionic (mABA) and zwitterionic (mABA$^{\pm}$) forms of meta-aminobenzoic acid in water were investigated by means of atomistic simulations. Density functional theory calculations with a polarisable continuum model to describe the aqueous environment were used to locate the low-lying energy structures and thermodynamic stability in water

of nonionic, $(mABA)_2$, zwitterionic, $(mABA^{\pm})_2$ and nonionic-zwitterionic, $(mABA)(mABA^{\pm})$, dimers. Results show that the only thermodynamically dimers in solution are $(mABA^{\pm})_2$, whereas the formation of the nonionic classic carboxylic dimer $(mABA)_2$ and the $\pi-\pi$ stacked $(mABA)(mABA^{\pm})$ dimer is endergonic. However, the free energies of $(mABA)_2$ formation are close to zero, and at high supersaturated conditions, such as the aqueous solutions considered in the MD simulations reported in Chapter 4, mABA molecules tend to aggregate spontaneously. Classical molecular dynamics simulations of meta-aminobenzoic acid aqueous solutions containing an equal amount of nonionic and zwitterionic species were conducted to examine the aggregation behavior as a function of concentration of solute. Analysis of the aggregates formed during the simulation shows a higher proportion of $\pi...\pi$ and $NH_3^+...COO^-$ pairs, which interactions occur in the most stable zwitterionic dimers $(mABA^{\pm})_2$ located using DFT calculations. According to the simulations reported in this work, the higher stability of zwitterionic dimers in solution drives the selective crystallisation of the zwitterionic polymorphs of meta-aminobenzoic acid. This work represents a paradigm of the role of molecular processes during the early stages of crystal nucleation in affecting polymorph selection during crystallisation from solution.

Chapter 6

Surface stability and reactivity of form-II of meta-aminobenzoic acid

Solution crystallisation usually occurs in the presence of a crystal seed, where the interactions between the surface and the solvent molecules at the solid-liquid interface play a primary role. In this study, we report atomistic simulations of the surface of form-II of meta-aminobenzoic acid (mABA) to gain a fundamental understanding of the role of solvent in controlling the seeded growth of mABA. Static density functional theory and *ab initio* molecular dynamics have been used to determine the stability of low-index surfaces and the adsorption of mABA units (monomer, dimer, tetramer). Classical molecular dynamics simulations of stable mABA surfaces in contact with water and dimethyl sulfoxide were then conducted to characterise the interactions occurring at the solid-liquid interface, and to quantify the time required to strip the overlying solvent molecules prior to formation

of the mABA growth units. The rate of solvent molecules removal at the interface is significantly dependent on the nature of the solvent and the crystallographic layer in contact with the solvent. Our results agree therefore with the surface inhibition hypothesis,[165] according to which solvent molecules interact differently with the surfaces of the molecular crystal and the rate of solvent removal can become an important rate-determining step in the growth of a given face and significantly influence the preferential growth of a specific polymorph.

6.1 Introduction

The nature of the solvent (polarity and proticity) or the presence in solution of additives (ions or molecules that are different from the constituents of the solvent and the crystal) can control the polymorphism of molecular crystals.[93,116,135] The molecule meta-aminobenzoic acid (mABA) can exist in several polymorphs whose nucleation depends chiefly on the solvent: polar and protic solvents such as water or methanol promote the crystallisation of the stable form I of mABA; aprotic solvents such as acetonitrile and dimethyl sulfoxide (DMSO) promote the formation of the metastable form II.[99] Another example is glycine, which has three polymorphs (α, β and γ): crystallisation from neutral aqueous solution leads to the formation of the α-form; β-glycine crystallises from water-alcohol solutions; the γ-form is promoted by the addition to the solution of sodium salts (NaCl, NaF and $NaNO_3$).[166] Changing the chemistry of the solution is a strategy routinely used in crystallisation experiments for the selection of polymorph

but it is still, largely, empirical, non-predictive, and heavily dependent on trial-and-error.

It is therefore of considerable technological importance to determine, at the fundamental level, how the processes surrounding the nucleation and crystal growth are influenced by the chemistry of solution, including the separation of solute and solvent molecules, the formation of structured molecular clusters in solution and the interaction at the solid-solution interface. Owing to advances in molecular dynamics (MD) techniques, as well as the availability of high-performance computing, it is now possible to simulate processes of crystallisation.[167–169] MD simulations of homogeneous solutions have shed light on the mechanisms controlling the early stages of crystal nucleation. Using a density functional theory (DFT) approach we proved that the most stable dimers and tetramers of mABA in solution correspond to the classic carboxylic dimer $\pi - \pi$ stacking synthon found in the crystalline form-II of mABA.[157,170] This result agrees with the "link hypothesis", according to which a supersaturated solution phase is populated by clusters of molecules having the packing of all the polymorphs of a molecular crystal, and the polymorph obtained from crystallisation is directly linked to the structure of the most abundant pre-nucleation molecular aggregates in solution.[75] Changes in the solution composition can therefore influence the relative stability of the crystal growth units and determine the nucleation of different polymorphs. However, extensive calculations conducted by Di Tommaso have questioned the validity of the link hypothesis by showing that the structure of the most thermodynamically stable species in a particular solvent is not always linked to the structural synthon of the polymorph that crystallises from that solution.[82] Another computational

study by Hamad *et al.* also suggested that the glycine monomer, and not the dimer, is the most abundant species in glycine aqueous solution over a wide concentration range.[87] The formation of a specific polymorph could therefore not only be depend on the population statistics of molecular clusters in solution.

Since solution crystallisation usually occurs in the presence of a crystal seed, the interactions at the solid-liquid interface between the surface and solvent molecules can therefore play a primary role in directing the polymorphic outcome of a crystallisation process. According to the "surface inhibition" hypothesis, solvent molecules interact differently with the surfaces of the molecular crystal and the removal of solvent can become a rate-determining step in the growth of a given face, significantly influencing the preferential growth of a specific polymorph. In this study, we report atomistic simulations of the solid-solution interface of form-II of meta-aminobenzoic acid. The aim is to gain a fundamental understanding of the role of solvent in controlling the seeded growth of mABA. Form II was considered because it is the most stable non-zwitterionic solid phase of mABA and selectively crystallises from DMSO solutions.[95]

Static DFT and *ab initio* MD simulations (AIMD) simulations have been used to determine the stability surfaces of form II of mABA and the energetics of adsorption of mABA units (monomer, dimer, tetramer) with the stable (001) and (010) surfaces. The (001) and (010) in contact with water and DMSO have been simulated with classical molecular dynamics (CMD) to characterise the interaction at the solid-liquid interface, and to quantify the time required to strip the overlying solvent molecules prior to the formation of the mABA growth units.

6.2 Computational methods

6.2.1 Surface construction

Starting from the experimental bulk unit cell of the form II of mABA,[171] the Materials Studio (version 7) software was used to generate seven low-index surfaces from the combination of the following Miller indices [hkl] (h, k, l = 0-1). To identify the stable surfaces of form II, different molecular models of the low-index surfaces of mABA were generated in which the size of the vacuum separating the two repeating surfaces in the supercell was kept constant at 20 Å, while the number of mABA layers composing the surface was varied to determine the convergence of the surface energy with respect to the thickness of the crystal slab.

6.2.2 Density functional theory calculations

Static DFT calculations were conducted with the periodic CASTEP (version 7.3)[172] and CP2K (version 2.5.1)[173] codes. CASTEP implements DFT using a plane-wave (PW) basis set and pseudopotentials approach. We have used the Perdew-Burke-Ernzerhof (PBE)[174] generalised gradient approximation for the exchange and correlation terms, with the Grimme-3 dispersion correction termed DFT-D3.[175] Vanderbilt ultrasoft pseudopotentials (USPP) represented core-valence interactions for all atomic species. Plane-wave basis set cutoffs for the wavefunctions were set to 50 and 500 Ry respectively. The Brillouin zone was sampled using 5x5x5 grids of k points following the Monkhorst-Pack scheme. CASTEP was used to fully optimise

the unit cell and obtain the equilibrium crystal structure. Lattice parameters and internal atomic coordinates were then independently modified to minimise the total energy and interatomic forces. The Broyden-Fletcher-Goldfarb-Shanno scheme was used for ions and cell relaxation. The criteria for the variable-cell minimisation were selected as follows: difference in total energy within 10^{-5} a.u., maximum ionic Hellmann-Feynman forces within 10-3 a.u., and maximum stress within 0.5 kbar. No symmetry constraints were applied in the calculation. Periodic boundary conditions were applied throughout.

CP2K implements DFT based on a hybrid Gaussian PW approach. We used PBE density functional together with DFT-D3. Goedecker-Teter-Hutter pseudopotentials[118] were used to describe the core-valence interactions. All atomic species were represented using a double-zeta valence polarized (DZVP) basis set with the PW kinetic energy cut-off was set to 1000 Ry. k-sampling was restricted to the Γ point of the Brillouin zone. Periodic boundary conditions were applied throughout. CP2K was used to conduct fixed-cell geometry optimisation of the mABA surfaces and to compute the adsorption energies of the monomer, dimer, tetramer on the stable (001) and (010) surfaces.

6.2.3 *Ab initio* molecular dynamics

Ab initio (Born-Oppenheimer) molecular dynamics (AIMD) simulations were conducted with the CP2K code. We used the PBE functional with the DFT-D3 dispersion correction, the Goedecker-Teter-Hutter pseudopotentials, and a plane wave kinetic energy cut of 1000 Ry. The Single-Zeta

Valence (SZV) polarised basis set was used for these simulations to allow for longer simulations times. The k-sampling was restricted to the Γ point of the Brillouin zone. Simulations were carried out with a wave function optimisation tolerance of 10^{-6} a.u. The time step was set to 1.0 fs. Periodic boundary conditions were applied throughout. AIMD simulations were carried out in the NVT ensemble using a Nosé-Hoover chain thermostat to maintain the average temperature at T = 300K. AIMD simulations were conducted to probe the potential energy surface (PES) associated with the adsorption of mABA units (monomer, dimer, tetramer) on the surface of the crystal.

6.2.4 Classical molecular dynamics

Forcefield. The generalised AMBER forcefield (GAFF)[141] was used to model the mABA surfaces and the DMSO molecules. The GAFF potential was previously used in our previous work[157] to characterise the aggregation from solution of mABA; this forcefield has been adopted to study the nucleation and crystal growth of other organic molecules[87,90,126,162,176] including the very similar para-aminobenzoic acid.[163,177] Water molecules were modelled using the SPC/E potential. The interaction between mABA surfaces and DMSO molecules and between mABA surfaces and water molecules were described using the GAFF potential. To obtain the forcefield parameters with the GAFF framework, structures and molecular electrostatic potentials (MEP) of mABA and DMSO were obtained using the Gaussian09 code. The Antechamber package was then used to compute partial charges according to the restrained electrostatic potential (RESP) formal-

ism.[143] The description of MEP with a selected quantum chemistry method and basis set is a key aspect in RESP derivation.[144] We used the HF/6-31G* method, which was the level of theory applied for the RESP charges derivation in Cornel *et al.* forcefield and successive modifications of the AMBER potential. This procedure gave us partial charges for DMSO consistent with Dupradeau *et al.*[144]

Simulation detail. Classical molecular dynamics (MD) simulations were conducted using version 5.0.4 of the GROMACS molecular dynamics package.[149] The leap-frog integrator algorithm with a time step of 2 fs was used to integrate the equation of motion. The isothermal-isobaric (constant NPT) ensemble was used to maintain a temperature of 300 K and a pressure of 1 bar. The velocity rescale thermostat and the isotropic Parrinello-Rahman barostat were used with 0.1 ps and 2.0 ps as the thermostat and barostat relaxation times respectively. The electrostatic forces were calculated by means of the particle-mesh Ewald approach with a cutoff of 1.0 nm. A 1.0 nm cutoff was also used for the van der Waals forces. The LINCS algorithm was applied at each step to preserve the bond lengths.

Simulation protocol. We performed classical MD simulations of mABA surfaces in contact with DMSO and water. Molecular models of solvated surfaces were generated using the *solvate* GROMACS utility to fill the empty simulation box with DMSO or water. Each solvated surface was then equilibrated using a three-step process. We minimised the solvated surface using the steepest descent algorithm with a tolerance on the maximum force of 300 $kJ.mol^{-1}.nm^{-1}$ before equilibrating the temperature and volume of each system by running 200 ps of constant volume, constant temperature (NVT) simulation followed by 100 ps of constant pressure,

constant temperature (NPT) simulation. Production runs in the NPT ensemble were then conducted for 20 sets of 10 ns each, for a total simulation time of 200 ns.

Analysis of simulations. Interatomic bonding pairs were analysed through the generation of radial distribution functions (RDF), $g_{\alpha\beta}(r)$ which represents the probability, relative to a random distribution, of finding atoms of types α and β separated by a distance r. The self-diffusion scalar coefficient of solvent molecules were calculated using the *msd* utility of GROMACS. The dynamics of the first solvation layer of the mABA surface were characterised using the direct method proposed by Hofer and co-workers,[178] which was successfully applied to quantify the water exchange frequency at structurally distinct sites on the calcium carbonate and calcium/magnesium carbonate surfaces,[179] and the dynamics of water molecules around hydrated alkaline earth metal ions and their carbonate and bicarbonate complexes.[30,180,181] To utilise this method, the trajectory files generated by GROMACS were processed using an in-house code named COMFORT (Centre Of Mass in FORTran), developed to generate trajectories of the center of masses (COMs) of the mABA and solvent (H_2O or DMSO) molecules from the all-atom trajectories generated using GROMACS. These MD trajectories were analysed for solvent molecule movements at the crystal-liquid interface and whenever the COM of a solvent molecule crossed the boundary of the coordination shell of mABA, its path was followed. If the new position outside or inside this shell lasted for more than a certain time parameter (1 ps $\leq \tau \leq$ 5 ps), then the event was counted as a real exchange event and used to quantify the reactivity of a surface in terms of the frequency of solvent exchange.[178] The coordination

shell of mABA was then defined by the first minimum of the COM(mABA)-COM(H_2O) and COM(mABA)-COM(DMSO) RDFs.

6.3 Stability of mABA surfaces

When modelling crystal growth, the solid-liquid interface of the most stable surface is usually considered.[90,179,182,183] However, experimental information related with the most dominant surface of form II and of the other polymorphs of mABA could not be found from literature data. The stability of low index planes of form II were therefore determined from DFT (PBE-D3) calculations of the surface energy (E_{surf}) according to the following equation:

$$E_{surf} = \frac{E_{slab,n} - n.E_{bulk}}{2.A_{surf}} \tag{6.1}$$

where E_{surf} is the surface energy, $E_{slab,n}$ is the energy of a slab of thickness n, E_{bulk} is the energy of the bulk and A_{surf} is the area of the base of the crystal.

For organic crystals, the number of mABA layers corresponds to the thickness parameter (Figure 6.1). Only those surfaces exhibiting convergence of the surface energy with respect to the number of layers are considered to be stable.

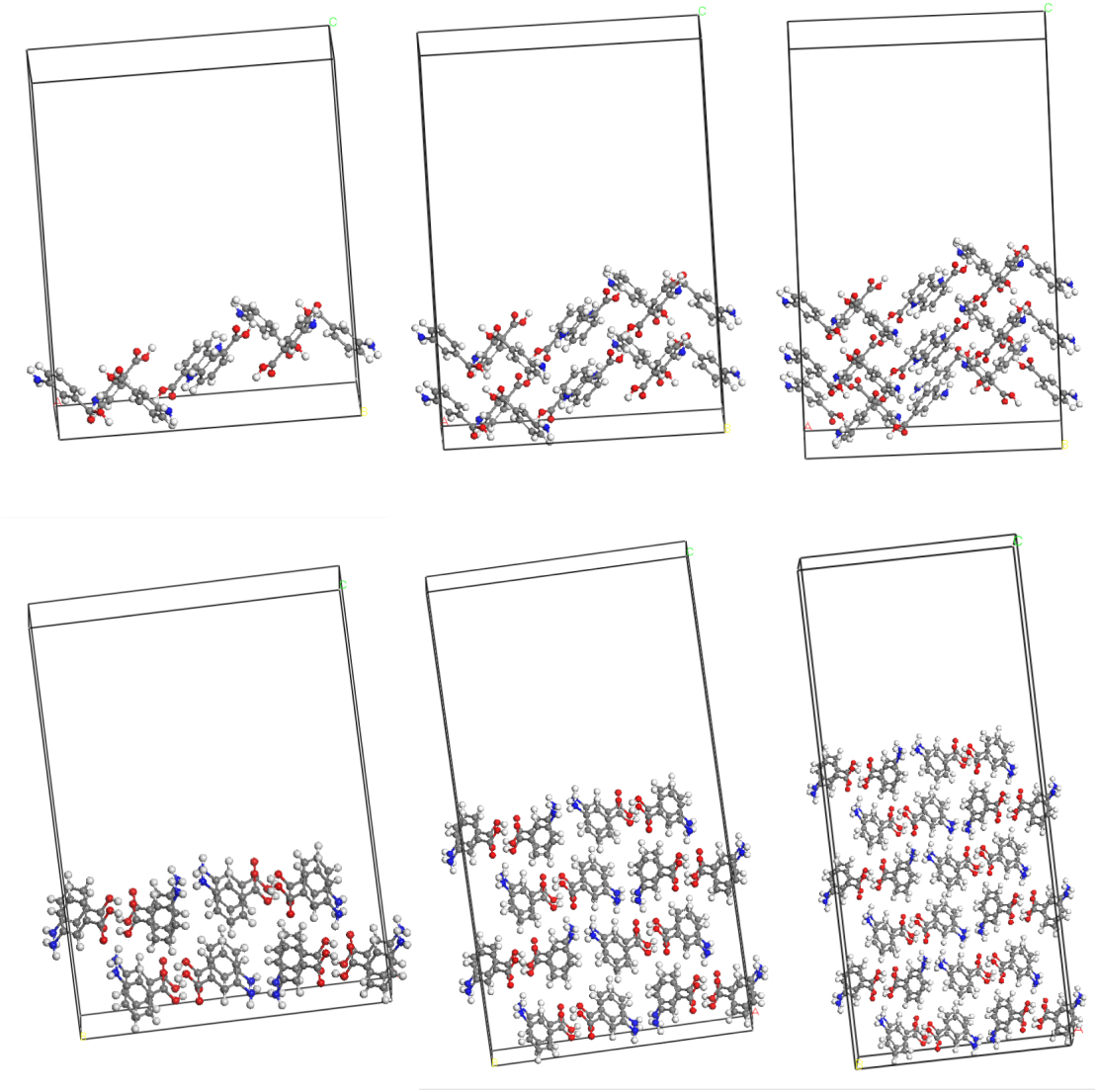

Figure 6.1: Molecular models of the unstable (100), (top) and (001), (bottom) surfaces of form II of mABA for different number of layers.

The evolution of the surface energies as a function of the thickness is reported in Figure 6.2. The convergent behavior of the surface energies of (001), (010) and (011) indicate that these faces can be considered stable. On the other hand, the surface energy of the other faces linear increases with the slab thickness and they are therefore unstable.

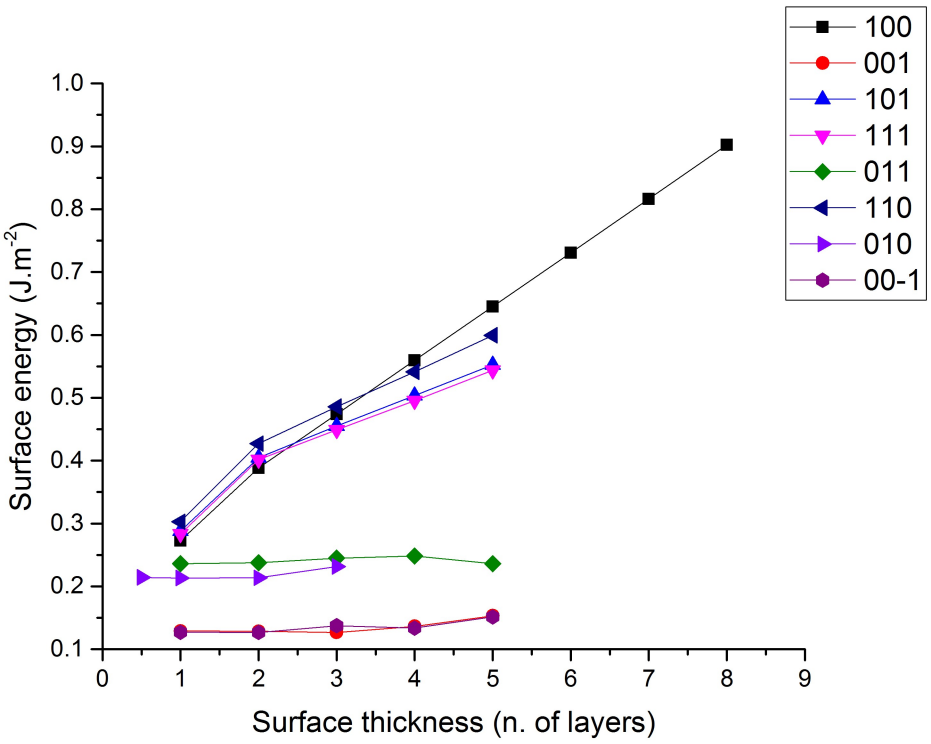

Figure 6.2: Evolution of the surface energies (in $J.m^{-2}$) with the number of layers in contact with a fixed 20 Å vacuum slab. Calculations conducted using the CP2K code with the PBE-D3 functional, the DZVP basis set and a PW cutoff of 1000 Ry. (001) and (00-1) are probably equivalent.

The stability and instability of the low index planes of form II can be explained using the concepts developed s by Bertaut and Tasker for ionic solids. Bertaut *et al.*[184] showed that if the crystalline structure is such that there is a dipole moment perpendicular to a surface, then the surface energy diverges and the surface is unstable. Tasker then demonstrated that, from a computational point of view, it is possible to model only those surfaces which crystallographic planes do not have a dipole moment that is approximately perpendicular to the surface.[185] Because the mABA molecule has a dipole moment (3.42 D at the M06-2X level of theory), then the optimised structures of the unstable surfaces (001) and (111) have an

overall non-zero dipole moment perpendicular to the surface (Figure 6.3). On the other hand, the arrangement of the mABA molecules in the crystallographic layers parallel to the (001) and (010) planes is such that there cannot be a dipole moment in the direction perpendicular to the surface and these surfaces will consequently be stable (Figure 6.3). Surfaces parallel to (001) and (010) can therefore be labelled of type 1 according to the Tasker classification.

The (001) and (010) faces were used to quantify the adsorption of mABA units with the crystal surface and to conduct MD simulations of the mABA crystal surface with water and DMSO.

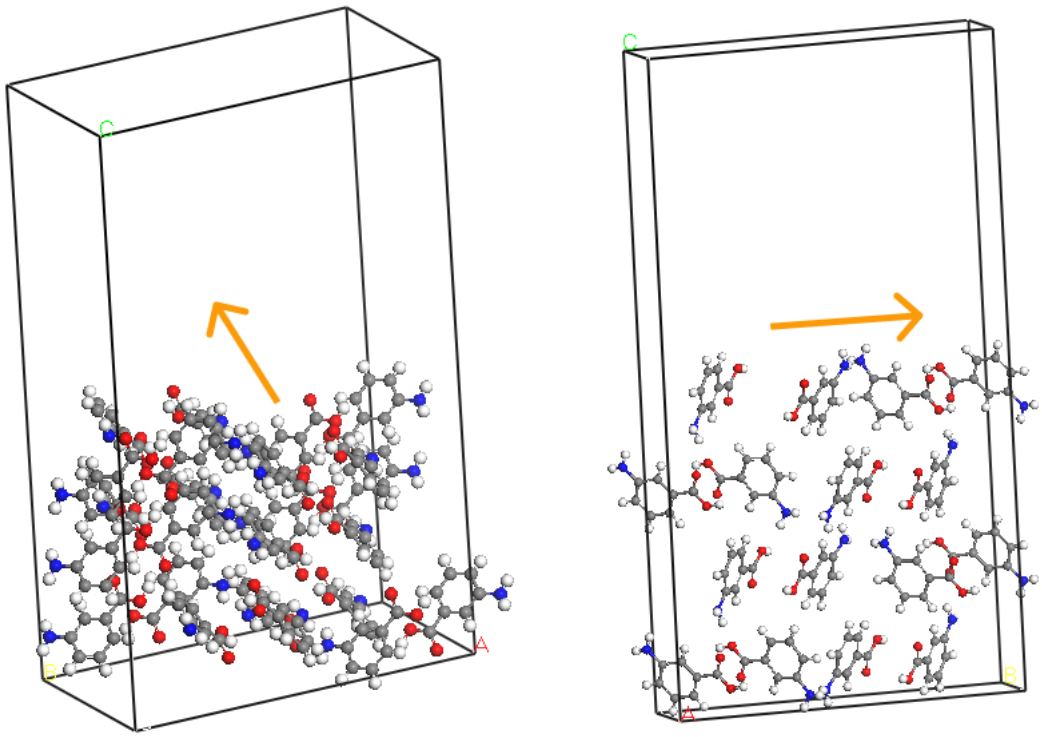

Figure 6.3: Simplified representation of the dipole moments in the unstable (100) surface (left) and stable (001) surface (right).

6.4 Adsorption of building units on stable mABA surfaces

The gas-phase binding energies for the adsorption of the monomer, dimer and tetramer of mABA to (001) and (010) surfaces were computed according to the following expression:

$$E_{bind} = E_{s+m} - (E_s + E_m) \tag{6.2}$$

where E_{bind} is the electronic binding energy of the surface with the molecular unit, $E_{(s+m)}$ is the total energy of the system (surface and molecular unit), E_s is the energy of the surface and E_m is the energy of the isolated molecular unit. Despite crystal growth occurring through the binding of solute molecules from the solution to the surface, gas-phase energetics can give an informative estimate regarding the preferential binding of the building units present in a mABA solution.

Our previous computational investigation of the aggregation of mABA from water and DMSO showed that the most stable dimers and tetramers in solution correspond to the classic carboxylic dimer π-π stacking synthon found in the crystalline form-II of mABA.[157] The monomer mABA, dimer $(mABA)_2$ and tetramer $(mABA)_4$ were placed on different locations of the (001) and (010) surfaces and subject to AIMD simulations for 20 ps to probe the PES of the system. The final configuration was then subject to full geometry optimisation. Table 6.1 reports the binding energies associated to the most stable adsorption sites of mABA, $(mABA)_2$ and $(mABA)_4$. The values indicate a generally favourable interaction between the mABA units

and the (001) surface but also suggest that the classic carboxylic dimer binds more strongly than monomers. The arrangement of these units in the optimised structures are reported in Figure 6.4. On the other hand, the adsorption of molecules on the (010) surface are highly exergonic.

Table 6.1: Binding energies (and binding energies per mABA unit) associated with the adsorption of monomer, dimer and tetramer on the (001) and (010) surfaces of form II of mABA. Values in kcal.mol^{-1}.

	E_{bind} (001)	E_{bind}/M (001)	E_{bind} (010)	E_{bind}/M (010)
Monomer, mABA	-12.24	-12.24	21.43	21.43
Dimer, $mABA_2$	-20.55	-10.27	-	-
Tetramer, $mABA_4$	-14.41	-3.60	-	-

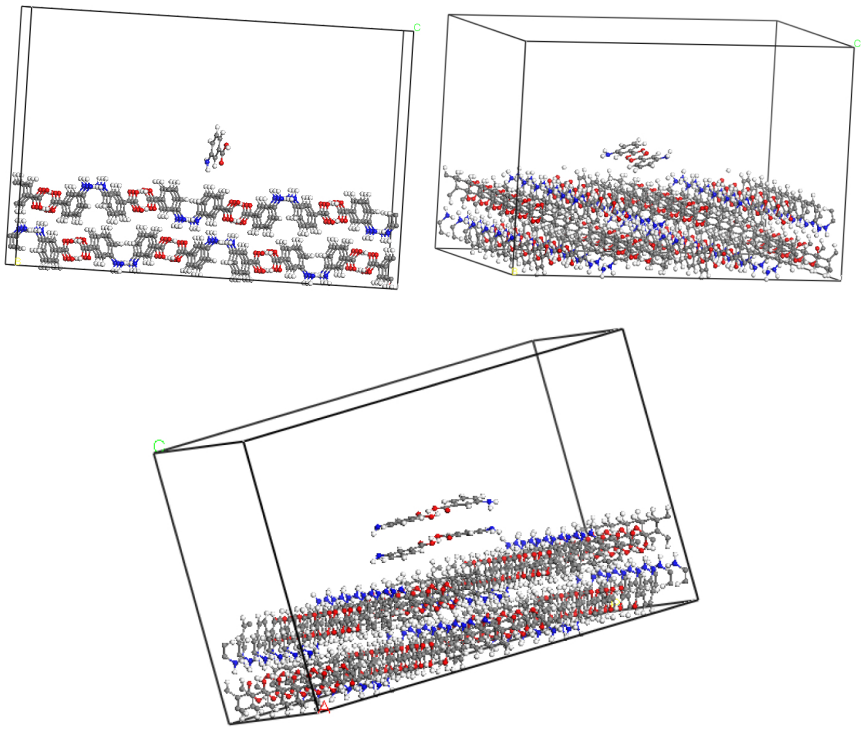

Figure 6.4: Arrangements of molecular units of mABA (monomer, dimer and tetramer) on top of the (001) surface of mABA form II.

6.5 Crystal-solvent interactions

The local interaction of the solvent with the mABA crystal was investigated by means of classical MD simulations by considering stable (001) and (010) surfaces in contact with DMSO and water. The (111) surface was also considered but after a short simulation period it underwent amorphisation, which confirmed the instability of this surface. Snapshots of the MD simulations of the (001) and (010) surfaces both in contact with water and DMSO are reported in Figure 6.5.

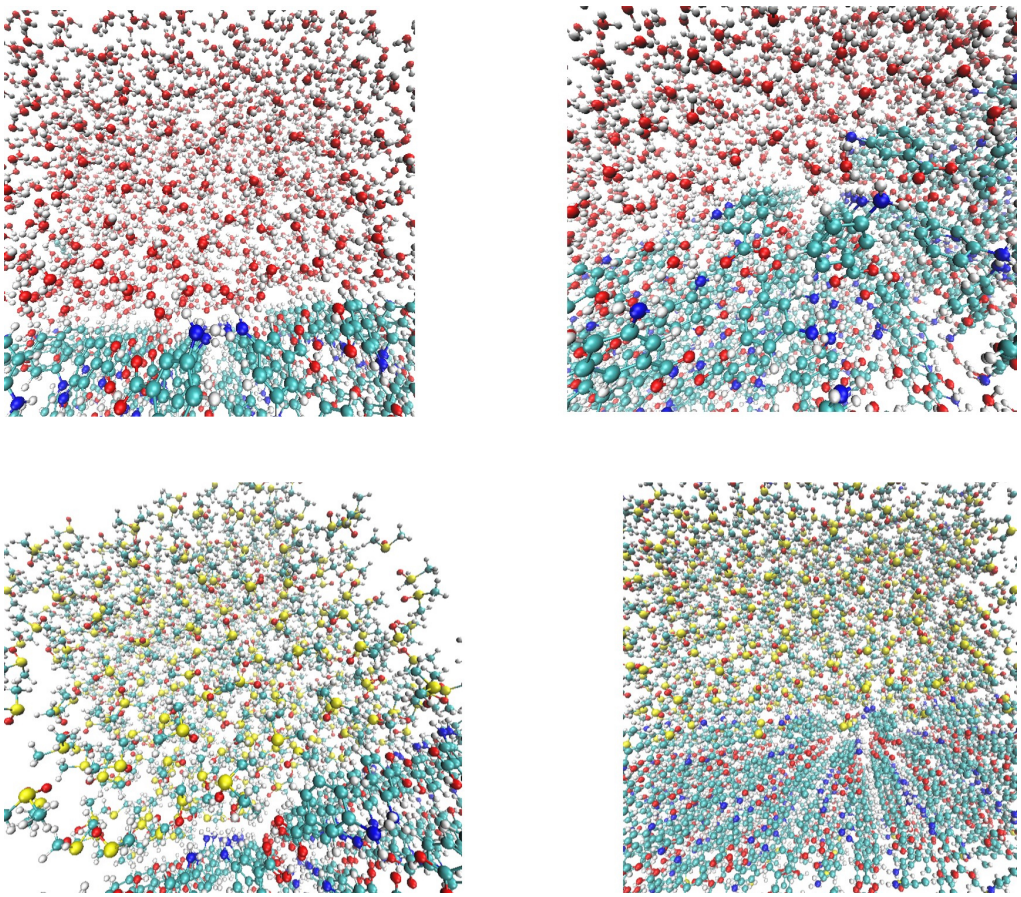

Figure 6.5: Final snapshots of molecular dynamics simulations for stable surfaces (001), *left* and (010), *right*, with both water, *top* and DMSO, *bottom*, at the solute-solvent interface.

The reactivity of the (001) and (010) surfaces was quantified in terms of the number of exchanges (N_{ex}) of solvent molecules at the solid-liquid interface using an algorithm based on the direct method proposed by Hofer *et al.*[186] The values of N_{ex} in Table 6.2 indicate a very different behavior of the (001) and (010) surfaces. The (010) surface does not display any solvent exchange event. As stripping the overlying solvent molecules is a slow process, the data in Table 6.2 are a clear indication of the low reactivity of the (010) surface. The removal of solvent molecules on the (001) surface is faster than on the (010) and is significantly dependent on the solvent

with four times higher number of solvent exchange events observed when the (001) surface is in contact with water. This is linked with our previous finding that the strength of mABA-DMSO interaction, measured in terms of solute-solvent binding free energy, is significantly larger (approximately 10 kcal.mol^{-1}) than mABA-water (Chapter 3). However, the comparison in Table 6.1 of the dynamics on the (001) and (010) surfaces suggest that the arrangement of mABA molecules in these two surfaces can play a significant role to the ordering of DMSO solvent molecules on the surface, and consequently on its dynamical properties. Table 6.3 finally reports a comparison of the number of exchange events for the mABA surface in contact with DMSO obtained using different time threshold of the τ parameter. As expected, an increase in τ decreases the observed number of exchange events but the trend with respect to the solvent is maintained.

Table 6.2: Number of exchanges (N_{ex}) of the solvent molecules at the (001) and (010) surfaces of form II of mABA. An exchange is considered to be real if the new position outside or inside the coordination shell of mABA lasts for more 0.5 ps. Values obtained from the analysis of the last 10 ns of every simulation period. Values normalised with respect to the number of mABA at the interface.

	001		010	
	DMSO	Water	DMSO	Water
50ns	140.5	2415.9	0.0	0.0
100ns	437.4	30.2	0.0	0.0
150ns	53.6	102.0	0.0	0.0
200ns	0.0	85.6	0.0	0.0

Table 6.3: Number of exchanges (N_{ex}) of the DMSO molecules at the (001) and (010) surfaces of form II of mABA. An exchange is considered to be real if the new position outside or inside the coordination shell of mABA lasts for more than τ. Values obtained from the analysis of the last 10 ns of every simulation period. Values normalised with respect to the number of mABA at the interface.

	$\tau = 1$ ps		$\tau = 3$ ps		$\tau = 5$ ps	
	(001)	(010)	(001)	(010)	(001)	(010)
50ns	87.1	0.0	29.7	0.0	21.8	0.0
100ns	235.5	0.0	69.3	0.0	35.6	0.0
150ns	25.8	0.0	7.9	0.0	7.9	0.0
200ns	0.0	0.0	0.0	0.0	0.0	0.0
Sum	348.4	0.0	106.9	0.0	65.3	0.0

6.6 Conclusions

In this chapter, the surface stability and reactivity of form-II of mABA in contact with water and DMSO have been investigated by means of atomistic simulations. Static DFT calculations have been used to compute the surface energy of the low index mABA surfaces and to rank the surfaces according to their stability. Classical MD simulations of the (001) and (010) stable surfaces in contact with liquid water and DMSO were then conducted to characterise the molecular interactions occurring at the solid-liquid interface and quantify the time needed to remove the solvent molecules on the mABA surface. This process is required before mABA growth units in solution (monomers or molecular aggregates) can adsorb on the surface. Our simulations indicate a strong dependence of the frequency of solvent removal from the solid-liquid interface with the nature of the solvent and the

topology of the surface. *Ab initio* MD simulations of the mABA monomer, dimer and tetramer on the (001) and (010) surfaces, followed by DFT optimisations showed a favourable interaction between mABA units and the (001) stable surface, while the adsorption of molecules on the (010) surface is highly exergonic. These simulations reveal the fundamental role of solvent and surface topology in controlling the growth of this important pharmaceutical product.

Chapter 7

Conclusions

This thesis developed and applied a combination of density functional theory (DFT) and molecular dynamics (MD) techniques to study the early stages of crystallisation of meta-aminobenzoic acid (mABA), an important model system in polymorphic research because of its ability to crystallise into five different crystal structures (I-V). The aim of this work was to identify what controls, at the molecular level, the selection process during crystallisation from solution of this important active pharmaceutical ingredient, widely used in the synthesis of analgesics, antihypertensives, vasodilators and other drugs.

Polymorph control during the formation of organic crystals from solution is influenced by the solute-solvent interactions, making solution environments highly influential on the molecular processes controlling crystallisation. Our study was split into three main areas: the solvation of mABA molecules, the self-association of mABA as a function of the solution environment, the stability and reactivity of mABA-solution interfaces.

We began by looking at the solvation of the monomers of meta-aminobenzoic acid in aqueous and organosulfur solutions. In aqueous solutions, meta-aminobenzoic has been reported 50% in the zwitterionic ($mABA^{\pm}$) and non-zwitterionic (mABA) forms; *ab initio* MD simulations were used to determine the stability and hydration hydration properties of these two species in water. A detailed analysis of mABA and $mABA^{\pm}$ with the surrounding water molecules shows that $mABA^{\pm}$-water interactions are stronger than mABA-water, and the interaction with the surrounding water molecules is stronger around the carboxylic acid than around the $-NH_2$ (mABA) and $-NH_3^+$ ($mABA^{\pm}$) groups. Although a coordination shell can be located for both monomers, the probability distributions of the number of water molecules surrounding mABA and $mABA^{\pm}$ show a great degree of flexibility in the hydration environment. A hybrid approach using a combination of explicit solvent molecules and a polarisable continuum model has then been used to quantify the equilibrium of the zwitterionic and non-zwitterionic species in organosulfur solutions, where the distribution of mABA and $mABA^{\pm}$ was still subject of debate. Our DFT calculations confirm that the free energy of reaction of the zwitterionic equilibrium mABA $\leftrightarrow$ $mABA^{\pm}$ is highly positive, supporting the hypothesis made by Hughes and co-workers that in dimethyl sulfoxide molecules of meta-aminobenzoic acid mainly exist in the non-zwitterionic form.[117] Further microsolvation DFT calculations were conducted to characterise the energetic barrier for the desolvation of mABA molecules and the strength of the mABA-solvent interactions; these are larger in dimethyl sulfoxide compared with water and suggest that the early stages of crystallisation of organic crystals could

be significantly influenced by the process of mABA desolvation.

In the subsequent part of the study, a DFT-based approach was developed to quantify the thermodynamics of meta-aminobenzoic acid aggregation from solution. In this methodology, the universal Solvation Model based on Density (SMD) is used to describe the solvation effects, and the low-lying energy structures of molecular complexes are located by adopting an efficient random-search procedure to probe the potential energy surfaces of mABA oligomers. Within this framework, the free energies of molecular clusters in solution are determined as the sum of the electronic energy, vibrational-rotational-translational gas-phase contribution and solvation free energy. Starting from randomly generated candidate structures and by imposing the condition of minimum free energy in solution for the isomers of $(\mathrm{mABA})_n$, we proved that the structure of the most stable dimer and tetramer correspond to the classic carboxylic dimer stacking synthon found in the crystalline form II of mABA. The transfer of structural information from the solution- to the solid-state phase of mABA are related with the presence of stable carboxylic acid dimers as well as higher-order $(\mathrm{mABA})_4$ prenucleation clusters. The presence in solution of these stable species could direct the nucleation process towards the formation of the polymorphic form II of mABA. MD simulations of mABA solutions at different concentrations were also conducted with the results showing a significant solvent-dependent aggregation behaviour of mABA: in water, mABA molecules form H-bonded π-π stacking molecular clusters; in organosulfur solutions, mABA molecules are in a more solvated state and high solute concentrations are required to observe appreciable levels of ag-

gregation. Moreover, the formation of the stable dimeric and tetrameric species in DMSO predicted by the DFT calculations could not be observed during the MD simulation. This apparently contradictory results were rationalised in terms of the free energy profiles for the dimerisation of mABA; the formation of $(mABA)_2$ in DMSO has an activation that is twice that in water and significantly larger than the value of kT at 300 K. This rationalises the observed solvent-dependent aggregation behaviour of mABA and agrees with the DFT calculations of mABA desolvation. This results again confirms the importance of specific mABA-solvent interactions during the initial stages of nucleation from solution.

In water, meta-aminobenzoic acid exists in both the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) forms. Therefore, we conducted further simulations to probe the events surrounding the early stage of crystal nucleation from aqueous solutions. DFT calculations show that the only thermodynamically dimers in solution are $(mABA^{\pm})_2$, whereas the formation of the nonionic classic carboxylic dimer $(mABA)_2$ and the π-π stacked $(mABA)(mABA^{\pm})$ dimer is endergonic. MD simulations of aqueous mixed mABA-$mABA^{\pm}$ were carried out to examine the aggregation of meta-aminobenzoic acid. Analysis of the aggregates formed during these simulations was conducted in terms of a symmetric pairwise interaction matrix to quantify the interactions between three body representations of the mABA and $mABA^{\pm}$. These analyses show a larger proportion of zwitterionic dimers. Therefore, the higher stability of zwitterionic dimers in aqueous solutions drives the selective crystallisation of the zwitterionic polymorphs of meta-aminobenzoic acid.

In the last part of this thesis, we presented preliminary work on the surface stability and reactivity of form-II of meta-aminobenzoic acid. Since solution crystallisation usually occurs in the presence of a crystal seed, we have conducted atomistic simulations to gain fundamental understanding of the role of solvent in controlling the seeded growth of mABA. Static DFT calculations have been used to determine the stability of low-index surfaces and the adsorption of mABA units (monomer, dimer, tetramer). Subsequently, MD simulations of the two stable mABA surfaces in contact with water and DMSO were conducted to characterise the interactions occurring at the solid-liquid interface and quantify the time required to strip the overlying solvent molecules prior to the surface adsorption of the mABA growth units. We found significant dependence of the frequency of solvent removal from the solid interface with the nature of the solvent and the crystallographic layer. This result agrees with the "surface inhibition hypothesis" proposed by Weissbuch and co-workers, according to which solvent molecules interact differently with the surfaces of the molecular crystal and the rate of solvent removal can become an important rate-determining step in the growth of a given face, significantly influencing the preferential growth of a specific polymorph.[165]

Within this study, we believe there is plenty of scope for future work. The current DFT-based methodology developed to simulate molecular aggregation is based on a random-searching procedure for the localisation of low-lying energy minima. Enhancements to this procedure could include the application of more sophisticated searching methods such as Basin-

Hopping or Genetic Algorithms. Moreover, the use of accurate interatomic potential forcefields to evaluate energies would substantially accelerate the searching process and allow the consideration of larger molecular clusters. The MD simulations of mABA solutions could be coupled with enhanced sampling methods such as metadynamics technique using the methodology recently proposed by Salvalaglio and co-workers.[90] The pairwise interaction matrix analysis could be applied to determine the collective variables used to drive the aggregation process during metadynamics. Regarding our preliminary study of surface growth, so far we have only considered the interaction with the pure solvent but future studies could include solutions containing varying concentrations of mABA as well as enhanced sampling methods (umbrella sampling and metadynamics) to compute the free energy profile associated with the adsorption of building units (monomer, dimer, tetramer) from the solution to the surface.

Appendix

A.1 Tables

Table A.1: Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.04 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage.

	A*	B*	C*	A	B	C
A*	0.2	0.5	7.7	2.0	1.9	2.7
B*		9.4	2.4	0.7	5.7	5.7
C*			0.1	1.9	0.5	3.9
A				6.5	4.7	7.7
B					11.8	15.7
C						8.5

Table A.2: Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.16 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage.

	A*	B*	C*	A	B	C
A*	0.2	0.9	9.3	4.1	3.3	3.6
B*		12.8	3.3	2.4	8.1	6.6
C*			0.3	3.7	1.9	3.0
A				5.0	3.7	6.4
B					6.6	9.1
C						5.6

Table A.3: Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.31 mol.L^{-1} mABA-mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage.

	A*	B*	C*	A	B	C
A*	0.3	0.7	9.3	4.6	3.5	3.9
B*		9.8	2.8	2.2	7.8	6.9
C*			0.3	3.7	1.7	3.8
A				5.9	4.0	6.3
B					6.5	9.7
C						6.6

A.2 GAFF parameters

mABA$^{\pm}$

```
; ZMB_GMX.top created by acpype (Rev: 403) on Sun Sep 4 21:51:21 2016

[ defaults ]
```

```
; nbfunc comb-rule gen-pairs fudgeLJ fudgeQQ
1 2 yes 0.5 0.8333
[ atomtypes ]
;name bond_type mass charge ptype sigma epsilon Amb
ca ca 0.00000 0.00000 A 3.39967e-01 3.59824e-01 ; 1.91 0.0860
ha ha 0.00000 0.00000 A 2.59964e-01 6.27600e-02 ; 1.46 0.0150
n4 n4 0.00000 0.00000 A 3.25000e-01 7.11280e-01 ; 1.82 0.1700
hn hn 0.00000 0.00000 A 1.06908e-01 6.56888e-02 ; 0.60 0.0157
c c 0.00000 0.00000 A 3.39967e-01 3.59824e-01 ; 1.91 0.0860
o o 0.00000 0.00000 A 2.95992e-01 8.78640e-01 ; 1.66 0.2100
[ moleculetype ]
;name nrexcl
ZMB 3
[ atoms ]
; nr type resi res atom cgnr charge mass ; qtot bond_type
1 ca 1 ZMB C2 1 -0.201672 12.01000 ; qtot -0.202
2 ha 1 ZMB H2 2 0.168307 1.00800 ; qtot -0.033
3 ca 1 ZMB C1 3 0.126418 12.01000 ; qtot 0.093
4 n4 1 ZMB N1 4 -0.389661 14.01000 ; qtot -0.297
5 hn 1 ZMB H5 5 0.335820 1.00800 ; qtot 0.039
6 hn 1 ZMB H6 6 0.335820 1.00800 ; qtot 0.375
7 hn 1 ZMB H7 7 0.335820 1.00800 ; qtot 0.711
8 ca 1 ZMB C6 8 -0.233164 12.01000 ; qtot 0.478
9 ha 1 ZMB H1 9 0.161128 1.00800 ; qtot 0.639
10 ca 1 ZMB C5 10 -0.204098 12.01000 ; qtot 0.435
11 ha 1 ZMB H4 11 0.174390 1.00800 ; qtot 0.609
```

```
12 ca 1 ZMB C4 12 -0.057139 12.01000 ; qtot 0.552
13 ha 1 ZMB H3 13 0.159004 1.00800 ; qtot 0.711
14 ca 1 ZMB C3 14 -0.022851 12.01000 ; qtot 0.688
15 c 1 ZMB C7 15 0.797939 12.01000 ; qtot 1.486
16 o 1 ZMB O2 16 -0.743031 16.00000 ; qtot 0.743
17 o 1 ZMB O1 17 -0.743031 16.00000 ; qtot 0.000
[ bonds ]
; ai aj funct r k
1 2 1 1.0870e-01 2.8811e+05 ; C2 - H2
1 3 1 1.3870e-01 4.0033e+05 ; C2 - C1
1 14 1 1.3870e-01 4.0033e+05 ; C2 - C3
3 4 1 1.4650e-01 2.7246e+05 ; C1 - N1
3 8 1 1.3870e-01 4.0033e+05 ; C1 - C6
4 5 1 1.0330e-01 3.0878e+05 ; N1 - H5
4 6 1 1.0330e-01 3.0878e+05 ; N1 - H6
4 7 1 1.0330e-01 3.0878e+05 ; N1 - H7
8 9 1 1.0870e-01 2.8811e+05 ; C6 - H1
8 10 1 1.3870e-01 4.0033e+05 ; C6 - C5
10 11 1 1.0870e-01 2.8811e+05 ; C5 - H4
10 12 1 1.3870e-01 4.0033e+05 ; C5 - C4
12 13 1 1.0870e-01 2.8811e+05 ; C4 - H3
12 14 1 1.3870e-01 4.0033e+05 ; C4 - C3
14 15 1 1.4870e-01 2.9263e+05 ; C3 - C7
15 16 1 1.2140e-01 5.4225e+05 ; C7 - O2
15 17 1 1.2140e-01 5.4225e+05 ; C7 - O1
[ pairs ]
```

```
; ai aj funct
1 5 1 ; C2 - H5
1 6 1 ; C2 - H6
1 7 1 ; C2 - H7
1 9 1 ; C2 - H1
1 10 1 ; C2 - C5
1 13 1 ; C2 - H3
1 16 1 ; C2 - O2
1 17 1 ; C2 - O1
2 4 1 ; H2 - N1
2 8 1 ; H2 - C6
2 12 1 ; H2 - C4
2 15 1 ; H2 - C7
3 11 1 ; C1 - H4
3 12 1 ; C1 - C4
3 15 1 ; C1 - C7
4 9 1 ; N1 - H1
4 10 1 ; N1 - C5
5 8 1 ; H5 - C6
6 8 1 ; H6 - C6
7 8 1 ; H7 - C6
8 13 1 ; C6 - H3
9 11 1 ; H1 - H4
9 12 1 ; H1 - C4
10 15 1 ; C5 - C7
11 13 1 ; H4 - H3
```

```
11 14 1 ; H4 - C3

12 16 1 ; C4 - O2

12 17 1 ; C4 - O1

13 15 1 ; H3 - C7

14 4 1 ; C3 - N1

14 8 1 ; C3 - C6

[ angles ]

; ai aj ak funct theta cth

1 3 4 1 1.1841e+02 5.6300e+02 ; C2 - C1 - N1

1 3 8 1 1.1997e+02 5.6216e+02 ; C2 - C1 - C6

1 14 12 1 1.1997e+02 5.6216e+02 ; C2 - C3 - C4

1 14 15 1 1.2014e+02 5.4091e+02 ; C2 - C3 - C7

2 1 3 1 1.2001e+02 4.0551e+02 ; H2 - C2 - C1

2 1 14 1 1.2001e+02 4.0551e+02 ; H2 - C2 - C3

3 1 14 1 1.1997e+02 5.6216e+02 ; C1 - C2 - C3

3 4 5 1 1.0852e+02 3.9781e+02 ; C1 - N1 - H5

3 4 6 1 1.0852e+02 3.9781e+02 ; C1 - N1 - H6

3 4 7 1 1.0852e+02 3.9781e+02 ; C1 - N1 - H7

3 8 9 1 1.2001e+02 4.0551e+02 ; C1 - C6 - H1

3 8 10 1 1.1997e+02 5.6216e+02 ; C1 - C6 - C5

4 3 8 1 1.1841e+02 5.6300e+02 ; N1 - C1 - C6

5 4 6 1 1.0811e+02 3.3907e+02 ; H5 - N1 - H6

5 4 7 1 1.0811e+02 3.3907e+02 ; H5 - N1 - H7

6 4 7 1 1.0811e+02 3.3907e+02 ; H6 - N1 - H7

8 10 11 1 1.2001e+02 4.0551e+02 ; C6 - C5 - H4

8 10 12 1 1.1997e+02 5.6216e+02 ; C6 - C5 - C4
```

```
9 8 10 1 1.2001e+02 4.0551e+02 ; H1 - C6 - C5
10 12 13 1 1.2001e+02 4.0551e+02 ; C5 - C4 - H3
10 12 14 1 1.1997e+02 5.6216e+02 ; C5 - C4 - C3
11 10 12 1 1.2001e+02 4.0551e+02 ; H4 - C5 - C4
12 14 15 1 1.2014e+02 5.4091e+02 ; C4 - C3 - C7
13 12 14 1 1.2001e+02 4.0551e+02 ; H3 - C4 - C3
14 15 16 1 1.2344e+02 5.7463e+02 ; C3 - C7 - O2
14 15 17 1 1.2344e+02 5.7463e+02 ; C3 - C7 - O1
16 15 17 1 1.3038e+02 6.5413e+02 ; O2 - C7 - O1
[ dihedrals ] ; propers
; treated as RBs in GROMACS to use combine multiple AMBER torsions per quartet
; i j k l func C0 C1 C2 C3 C4 C5
1 3 4 5 3 0.00000 0.00000 14.64400 0.00000 0.00000
0.00000 ; C2- C1- N1- H5
1 3 4 6 3 0.00000 0.00000 14.64400 0.00000 0.00000
0.00000 ; C2- C1- N1- H6
1 3 4 7 3 0.00000 0.00000 14.64400 0.00000 0.00000
0.00000 ; C2- C1- N1- H7
1 3 8 9 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C1- C6- H1
1 3 8 10 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C1- C6- C5
1 14 12 10 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C3- C4- C5
1 14 12 13 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C3- C4- H3
```

```
1 14 15 16 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C2- C3- C7- O2
1 14 15 17 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C2- C3- C7- O1
2 1 3 4 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C1- N1
2 1 3 8 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C1- C6
2 1 14 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C3- C4
2 1 14 15 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C3- C7
3 1 14 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C2- C3- C4
3 1 14 15 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C2- C3- C7
3 8 10 11 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C6- C5- H4
3 8 10 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C6- C5- C4
4 3 8 9 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; N1- C1- C6- H1
4 3 8 10 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; N1- C1- C6- C5
5 4 3 8 3 0.00000 0.00000 14.64400 0.00000 0.00000
0.00000 ; H5- N1- C1- C6
```

```
6 4 3 8 3 0.00000 0.00000 14.64400 0.00000 0.00000
0.00000 ; H6- N1- C1- C6
7 4 3 8 3 0.00000 0.00000 14.64400 0.00000 0.00000
0.00000 ; H7- N1- C1- C6
8 10 12 13 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C6- C5- C4- H3
8 10 12 14 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C6- C5- C4- C3
9 8 10 11 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H1- C6- C5- H4
9 8 10 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H1- C6- C5- C4
10 12 14 15 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C5- C4- C3- C7
11 10 12 13 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H4- C5- C4- H3
11 10 12 14 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H4- C5- C4- C3
12 14 15 16 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C4- C3- C7- O2
12 14 15 17 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C4- C3- C7- O1
13 12 14 15 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H3- C4- C3- C7
14 1 3 4 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C3- C2- C1- N1
```

```
14 1 3 8 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C3- C2- C1- C6
[ dihedrals ] ; impropers
; treated as propers in GROMACS to use correct AMBER analytical function
; i j k l func phase kd pn
1 8 3 4 1 180.00 4.60240 2 ; C2- C6- C1- N1
2 1 14 3 1 180.00 4.60240 2 ; H2- C2- C3- C1
3 10 8 9 1 180.00 4.60240 2 ; C1- C5- C6- H1
8 12 10 11 1 180.00 4.60240 2 ; C6- C4- C5- H4
10 14 12 13 1 180.00 4.60240 2 ; C5- C3- C4- H3
14 16 15 17 1 180.00 4.60240 2 ; C3- O2- C7- O1
15 1 14 12 1 180.00 4.60240 2 ; C7- C2- C3- C4
[ system ]
ZMB
[ molecules ]
; Compound nmols
ZMB 1
```

mABA

```
; MBA_GMX.top created by acpype (Rev: 403) on Thu Jul 9 22:34:48 2015
[ defaults ]
; nbfunc comb-rule gen-pairs fudgeLJ fudgeQQ
1 2 yes 0.5 0.8333
[ atomtypes ]
;name bond_type mass charge ptype sigma epsilon Amb
 ca ca 0.00000 0.00000 A 3.39967e-01 3.59824e-01 ; 1.91 0.0860
 ha ha 0.00000 0.00000 A 2.59964e-01 6.27600e-02 ; 1.46 0.0150
```

```
 nh nh 0.00000 0.00000 A 3.25000e-01 7.11280e-01 ; 1.82 0.1700
 hn hn 0.00000 0.00000 A 1.06908e-01 6.56888e-02 ; 0.60 0.0157
 c c 0.00000 0.00000 A 3.39967e-01 3.59824e-01 ; 1.91 0.0860
 oh oh 0.00000 0.00000 A 3.06647e-01 8.80314e-01 ; 1.72 0.2104
 ho ho 0.00000 0.00000 A 0.00000e+00 0.00000e+00 ; 0.00 0.0000
 o o 0.00000 0.00000 A 2.95992e-01 8.78640e-01 ; 1.66 0.2100
[ moleculetype ]
;name nrexcl
 MBA 3
[ atoms ]
; nr type resi res atom cgnr charge mass ; qtot bond_type
 1 ca 1 MBA C2 1 -0.185783 12.01000 ; qtot -0.186
 2 ha 1 MBA H2 2 0.172356 1.00800 ; qtot -0.013
 3 ca 1 MBA C1 3 0.297382 12.01000 ; qtot 0.284
 4 nh 1 MBA N1 4 -0.890198 14.01000 ; qtot -0.606
 5 hn 1 MBA H5 5 0.379769 1.00800 ; qtot -0.226
 6 hn 1 MBA H6 6 0.379769 1.00800 ; qtot 0.153
 7 ca 1 MBA C6 7 -0.178160 12.01000 ; qtot -0.025
 8 ha 1 MBA H1 8 0.160780 1.00800 ; qtot 0.136
 9 ca 1 MBA C5 9 -0.218354 12.01000 ; qtot -0.082
 10 ha 1 MBA H4 10 0.169120 1.00800 ; qtot 0.087
 11 ca 1 MBA C4 11 -0.118964 12.01000 ; qtot -0.032
 12 ha 1 MBA H3 12 0.157379 1.00800 ; qtot 0.125
 13 ca 1 MBA C3 13 -0.112371 12.01000 ; qtot 0.013
 14 c 1 MBA C7 14 0.773096 12.01000 ; qtot 0.786
 15 oh 1 MBA O2 15 -0.641122 16.00000 ; qtot 0.145
```

```
 16 ho 1 MBA H7 16 0.451189 1.00800 ; qtot 0.596
 17 o 1 MBA O1 17 -0.595890 16.00000 ; qtot 0.000
[ bonds ]
; ai aj funct r k
 1 2 1 1.0870e-01 2.8811e+05 ; C2 - H2
 1 3 1 1.3870e-01 4.0033e+05 ; C2 - C1
 1 13 1 1.3870e-01 4.0033e+05 ; C2 - C3
 3 4 1 1.3640e-01 3.7572e+05 ; C1 - N1
 3 7 1 1.3870e-01 4.0033e+05 ; C1 - C6
 4 5 1 1.0140e-01 3.3572e+05 ; N1 - H5
 4 6 1 1.0140e-01 3.3572e+05 ; N1 - H6
 7 8 1 1.0870e-01 2.8811e+05 ; C6 - H1
 7 9 1 1.3870e-01 4.0033e+05 ; C6 - C5
 9 10 1 1.0870e-01 2.8811e+05 ; C5 - H4
 9 11 1 1.3870e-01 4.0033e+05 ; C5 - C4
 11 12 1 1.0870e-01 2.8811e+05 ; C4 - H3
 11 13 1 1.3870e-01 4.0033e+05 ; C4 - C3
 13 14 1 1.4870e-01 2.9263e+05 ; C3 - C7
 14 15 1 1.3060e-01 3.9028e+05 ; C7 - O2
 14 17 1 1.2140e-01 5.4225e+05 ; C7 - O1
 15 16 1 9.7400e-02 3.0928e+05 ; O2 - H7
[ pairs ]
; ai aj funct
 1 5 1 ; C2 - H5
 1 6 1 ; C2 - H6
 1 8 1 ; C2 - H1
```

```
1 9 1 ; C2 - C5

1 12 1 ; C2 - H3

1 15 1 ; C2 - O2

1 17 1 ; C2 - O1

2 4 1 ; H2 - N1

2 7 1 ; H2 - C6

2 11 1 ; H2 - C4

2 14 1 ; H2 - C7

3 10 1 ; C1 - H4

3 11 1 ; C1 - C4

3 14 1 ; C1 - C7

4 8 1 ; N1 - H1

4 9 1 ; N1 - C5

5 7 1 ; H5 - C6

6 7 1 ; H6 - C6

7 12 1 ; C6 - H3

8 10 1 ; H1 - H4

8 11 1 ; H1 - C4

9 14 1 ; C5 - C7

10 12 1 ; H4 - H3

10 13 1 ; H4 - C3

11 15 1 ; C4 - O2

11 17 1 ; C4 - O1

12 14 1 ; H3 - C7

13 4 1 ; C3 - N1

13 7 1 ; C3 - C6
```

```
 13 16 1 ; C3 - H7
 16 17 1 ; H7 - O1
[ angles ]
; ai aj ak funct theta cth
 1 3 4 1 1.2013e+02 5.8024e+02 ; C2 - C1 - N1
 1 3 7 1 1.1997e+02 5.6216e+02 ; C2 - C1 - C6
 1 13 11 1 1.1997e+02 5.6216e+02 ; C2 - C3 - C4
 1 13 14 1 1.2014e+02 5.4091e+02 ; C2 - C3 - C7
 2 1 3 1 1.2001e+02 4.0551e+02 ; H2 - C2 - C1
 2 1 13 1 1.2001e+02 4.0551e+02 ; H2 - C2 - C3
 3 1 13 1 1.1997e+02 5.6216e+02 ; C1 - C2 - C3
 3 4 5 1 1.1613e+02 4.1070e+02 ; C1 - N1 - H5
 3 4 6 1 1.1613e+02 4.1070e+02 ; C1 - N1 - H6
 3 7 8 1 1.2001e+02 4.0551e+02 ; C1 - C6 - H1
 3 7 9 1 1.1997e+02 5.6216e+02 ; C1 - C6 - C5
 4 3 7 1 1.2013e+02 5.8024e+02 ; N1 - C1 - C6
 5 4 6 1 1.1485e+02 3.3514e+02 ; H5 - N1 - H6
 7 9 10 1 1.2001e+02 4.0551e+02 ; C6 - C5 - H4
 7 9 11 1 1.1997e+02 5.6216e+02 ; C6 - C5 - C4
 8 7 9 1 1.2001e+02 4.0551e+02 ; H1 - C6 - C5
 9 11 12 1 1.2001e+02 4.0551e+02 ; C5 - C4 - H3
 9 11 13 1 1.1997e+02 5.6216e+02 ; C5 - C4 - C3
 10 9 11 1 1.2001e+02 4.0551e+02 ; H4 - C5 - C4
 11 13 14 1 1.2014e+02 5.4091e+02 ; C4 - C3 - C7
 12 11 13 1 1.2001e+02 4.0551e+02 ; H3 - C4 - C3
 13 14 15 1 1.1344e+02 5.8668e+02 ; C3 - C7 - O2
```

```
 13 14 17 1 1.2344e+02 5.7463e+02 ; C3 - C7 - O1
 14 15 16 1 1.0737e+02 4.2836e+02 ; C7 - O2 - H7
 15 14 17 1 1.2288e+02 6.4752e+02 ; O2 - C7 - O1
[ dihedrals ] ; propers
; treated as RBs in GROMACS to use combine multiple AMBER torsions per quartet
; i j k l func C0 C1 C2 C3 C4 C5
 1 3 4 5 3 8.78640 0.00000 -8.78640 0.00000 0.00000
0.00000 ; C2- C1- N1- H5
 1 3 4 6 3 8.78640 0.00000 -8.78640 0.00000 0.00000
0.00000 ; C2- C1- N1- H6
 1 3 7 8 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C1- C6- H1
 1 3 7 9 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C1- C6- C5
 1 13 11 9 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C3- C4- C5
 1 13 11 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C2- C3- C4- H3
 1 13 14 15 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C2- C3- C7- O2
 1 13 14 17 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C2- C3- C7- O1
 2 1 3 4 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C1- N1
 2 1 3 7 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C1- C6
```

```
 2 1 13 11 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C3- C4
 2 1 13 14 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H2- C2- C3- C7
 3 1 13 11 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C2- C3- C4
 3 1 13 14 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C2- C3- C7
 3 7 9 10 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C6- C5- H4
 3 7 9 11 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C1- C6- C5- C4
 4 3 7 8 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; N1- C1- C6- H1
 4 3 7 9 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; N1- C1- C6- C5
 5 4 3 7 3 8.78640 0.00000 -8.78640 0.00000 0.00000
0.00000 ; H5- N1- C1- C6
 6 4 3 7 3 8.78640 0.00000 -8.78640 0.00000 0.00000
0.00000 ; H6- N1- C1- C6
 7 9 11 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C6- C5- C4- H3
 7 9 11 13 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C6- C5- C4- C3
 8 7 9 10 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H1- C6- C5- H4
```

```
 8 7 9 11 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H1- C6- C5- C4
 9 11 13 14 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C5- C4- C3- C7
 10 9 11 12 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H4- C5- C4- H3
 10 9 11 13 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H4- C5- C4- C3
 11 13 14 15 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C4- C3- C7- O2
 11 13 14 17 3 8.36800 0.00000 -8.36800 0.00000 0.00000
0.00000 ; C4- C3- C7- O1
 12 11 13 14 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; H3- C4- C3- C7
 13 1 3 4 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C3- C2- C1- N1
 13 1 3 7 3 30.33400 0.00000 -30.33400 0.00000 0.00000
0.00000 ; C3- C2- C1- C6
 13 14 15 16 3 19.24640 0.00000 -19.24640 0.00000 0.00000
0.00000 ; C3- C7- O2- H7
 16 15 14 17 3 27.19600 -7.94960 -19.24640 0.00000 0.00000
0.00000 ; H7- O2- C7- O1
[ dihedrals ] ; impropers
; treated as propers in GROMACS to use correct AMBER analytical function
; i j k l func phase kd pn
 1 7 3 4 1 180.00 4.60240 2 ; C2- C6- C1- N1
```

```
 2 1 13 3 1 180.00 4.60240 2 ; H2- C2- C3- C1
 3 5 4 6 1 180.00 4.60240 2 ; C1- H5- N1- H6
 3 9 7 8 1 180.00 4.60240 2 ; C1- C5- C6- H1
 7 11 9 10 1 180.00 4.60240 2 ; C6- C4- C5- H4
 9 13 11 12 1 180.00 4.60240 2 ; C5- C3- C4- H3
 13 17 14 15 1 180.00 43.93200 2 ; C3- O1- C7- O2
 14 1 13 11 1 180.00 4.60240 2 ; C7- C2- C3- C4
[ system ]
 MBA
[ molecules ]
; Compound nmols
 MBA 1
```

A.3 Codes in Fortran

A.3.1 COMFORT

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Etienne Gaines, Devis Di Tommaso
Queen Mary - University of London
Centre Of Mass in FORTran
November 2016
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

IMPLICIT NONE

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! Declaring the variables to be used in the program
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

```
CHARACTER(LEN=5)  :: atnam

CHARACTER(LEN=8)  :: res

CHARACTER(LEN=50) :: title

CHARACTER(LEN=30) :: FMT

CHARACTER(LEN=30) :: FMT2


INTEGER :: ierror_1, ierror_2

INTEGER :: pos

INTEGER :: MD_nsteps

INTEGER :: atms_tot

INTEGER :: MBA_nmol, DMS_nmol

INTEGER :: nmol

INTEGER :: MBA_natms, DMS_natms

INTEGER :: i, j, k

INTEGER, PARAMETER :: dbl = selected_real_kind(p=12)


REAL (KIND=dbl) :: m_atnam

REAL (KIND=dbl) :: M_MBA, M_DMS, m_H, m_C, m_N, m_O, m_S

REAL (KIND=dbl) :: M_MOL

REAL (KIND=dbl) :: xx, yy, zz

REAL (KIND=dbl) :: thetaX, thetaY, thetaZ, thetaBX, thetaBY, thetaBZ

REAL (KIND=dbl) :: UUXa, UUYa, UUZa, VVXa, VVYa, VVZa

REAL (KIND=dbl), ALLOCATABLE, DIMENSION(:) :: UUX1, UUY1, UUZ1

REAL (KIND=dbl), ALLOCATABLE, DIMENSION(:) :: VVX1, VVY1, VVZ1

REAL (KIND=dbl), ALLOCATABLE, DIMENSION(:) :: UUX2, UUY2, UUZ2

REAL (KIND=dbl), ALLOCATABLE, DIMENSION(:) :: VVX2, VVY2, VVZ2

REAL (KIND=dbl), ALLOCATABLE, DIMENSION(:) :: XXF1, YYF1, ZZF1

REAL (KIND=dbl), ALLOCATABLE, DIMENSION(:) :: XXF2, YYF2, ZZF2

REAL (KIND=dbl) :: AX, AY, AZ, dummy

REAL (KIND=dbl) :: PI
```

```
FMT = "(3f8.5)"

FMT2 = "(a7,i8,3f12.5)"

dummy = 0.

PI = 3.141592


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! Open the PARAM file that contains the values for the masses

! of the system and its constituants

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!


OPEN (UNIT=10, FILE='PARAM', STATUS='old', action='read', iostat=ierror_1)

IF (ierror_1 == 0) THEN

READ (10,*) MD_nsteps

READ (10,*) M_MBA

READ (10,*) M_DMS

READ (10,*) MBA_nmol

READ (10,*) DMS_nmol

READ (10,*) MBA_natms

READ (10,*) DMS_natms

READ (10,*) m_H

READ (10,*) m_C

READ (10,*) m_N

READ (10,*) m_O

READ (10,*) m_S


        ALLOCATE (UUX1(MBA_natms*MBA_nmol))

        ALLOCATE (UUY1(MBA_natms*MBA_nmol))

        ALLOCATE (UUZ1(MBA_natms*MBA_nmol))

        ALLOCATE (VVX1(MBA_natms*MBA_nmol))

        ALLOCATE (VVY1(MBA_natms*MBA_nmol))

        ALLOCATE (VVZ1(MBA_natms*MBA_nmol))
```

```
        ALLOCATE (XXF1(MBA_natms*MBA_nmol))
        ALLOCATE (YYF1(MBA_natms*MBA_nmol))
        ALLOCATE (ZZF1(MBA_natms*MBA_nmol))
        ALLOCATE (UUX2(DMS_natms*DMS_nmol))
        ALLOCATE (UUY2(DMS_natms*DMS_nmol))
        ALLOCATE (UUZ2(DMS_natms*DMS_nmol))
        ALLOCATE (VVX2(DMS_natms*DMS_nmol))
        ALLOCATE (VVY2(DMS_natms*DMS_nmol))
        ALLOCATE (VVZ2(DMS_natms*DMS_nmol))
        ALLOCATE (XXF2(DMS_natms*DMS_nmol))
        ALLOCATE (YYF2(DMS_natms*DMS_nmol))
        ALLOCATE (ZZF2(DMS_natms*DMS_nmol))

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! Open the GRO_FILE that contains the trajectories and coord.
! to compute the Center of Mass of each molecule
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

OPEN (UNIT=11, FILE='GRO_TRAJ', STATUS='old', action='read', iostat=ierror_2)

IF (ierror_2 == 0) THEN

DO i=1, MD_nsteps
READ (11,*) title
        READ (11,FMT) AX, AY, AZ

DO j=1, MBA_nmol

        UUXa = 0.
UUYa = 0.
UUZa = 0.
VVXa = 0.
```

```
VVYa = 0.

VVZa = 0.


DO k=1, MBA_natms

READ (11,40) nmol, res, atnam, xx, yy, zz

40 FORMAT (i7,a8,a5,3f12.5)


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! Assign the 'm_atnam' variable to the right value

! depending on the atom considered

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!


                IF (atnam(4:4) .eq. 'H') THEN

                       m_atnam = m_H

                ELSE IF (atnam(4:4) .eq. 'C') THEN

                        m_atnam = m_C

                ELSE IF (atnam(4:4) .eq. 'N') THEN

                        m_atnam = m_N

ELSE IF (atnam(4:4) .eq. 'O') THEN

                       m_atnam = m_O

ENDIF


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! Compute the Center of Mass for the X, Y and Z coord.

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!


thetaX=2*PI*xx/AX

thetaY=2*PI*yy/AY

thetaZ=2*PI*zz/AZ


UUX1(k)=cos(thetaX)

UUY1(k)=cos(thetaY)
```

```
UUZ1(k)=cos(thetaZ)

VVX1(k)=sin(thetaX)

VVY1(k)=sin(thetaY)

VVZ1(k)=sin(thetaZ)


UUXa = UUXa + m_atnam*UUX1(k)/M_MBA

UUYa = UUYa + m_atnam*UUY1(k)/M_MBA

UUZa = UUZa + m_atnam*UUZ1(k)/M_MBA

VVXa = VVXa + m_atnam*VVX1(k)/M_MBA

VVYa = VVYa + m_atnam*VVY1(k)/M_MBA

VVZa = VVZa + m_atnam*VVZ1(k)/M_MBA


thetaBX = atan2(-UUXa,-VVXa)+PI

thetaBY = atan2(-UUYa,-VVYa)+PI

thetaBZ = atan2(-UUZa,-VVZa)+PI


XXF1(j)=AX*thetaBX/(2*PI)

YYF1(j)=AY*thetaBY/(2*PI)

ZZF1(j)=AZ*thetaBZ/(2*PI)


ENDDO

END DO

DO j=1, DMS_nmol


        UUXa = 0.

UUYa = 0.

UUZa = 0.

VVXa = 0.

VVYa = 0.

VVZa = 0.


DO k=1, DMS_natms
```

```
                READ (11,40) nmol, res, atnam, xx, yy, zz


                IF (atnam(4:4) .eq. 'H') THEN
                        m_atnam = m_H
                ELSE IF (atnam(4:4) .eq. 'C') THEN
                        m_atnam = m_C
                ELSE IF (atnam(4:4) .eq. 'O') THEN
                        m_atnam = m_O
                ELSE IF (atnam(4:4) .eq. 'N') THEN
                        m_atnam = m_N
ELSE IF (atnam(4:4) .eq. 'S') THEN
                        m_atnam = m_S
                END IF


thetaX=2*PI*xx/AX
thetaY=2*PI*yy/AY
thetaZ=2*PI*zz/AZ


UUX2(k) = cos(thetaX)
UUY2(k) = cos(thetaY)
UUZ2(k) = cos(thetaZ)
VVX2(k) = sin(thetaX)
VVY2(k) = sin(thetaY)
VVZ2(k) = sin(thetaZ)


UUXa = UUXa + m_atnam*UUX2(k)/M_DMS
UUYa = UUYa + m_atnam*UUY2(k)/M_DMS
UUZa = UUZa + m_atnam*UUZ2(k)/M_DMS
VVXa = VVXa + m_atnam*VVX2(k)/M_DMS
VVYa = VVYa + m_atnam*VVY2(k)/M_DMS
VVZa = VVZa + m_atnam*VVZ2(k)/M_DMS
```

```
thetaBX = atan2(-UUXa,-VVXa)+PI

thetaBY = atan2(-UUYa,-VVYa)+PI

thetaBZ = atan2(-UUZa,-VVZa)+PI


XXF2(j) = AX*thetaBX/(2*PI)

YYF2(j) = AY*thetaBY/(2*PI)

ZZF2(j) = AZ*thetaBZ/(2*PI)


END DO


END DO


WRITE (*,*) 'Timestep = ', i

WRITE (*,FMT) AX, dummy, dummy

WRITE (*,FMT) dummy, AY, dummy

WRITE (*,FMT) dummy, dummy, AZ


DO j = 1, MBA_nmol

WRITE (*,FMT2) 'COM_MBA', j,  XXF1(j), YYF1(j), ZZF1(j)

ENDDO

DO j = 1, DMS_nmol

WRITE (*,FMT2) 'COM_DMS', j+MBA_nmol, XXF2(j), YYF2(j), ZZF2(j)

ENDDO


END DO

ELSE

        WRITE (*,*) " ierror_2 = ", ierror_2

ENDIF


DEALLOCATE (UUX1)

DEALLOCATE (UUY1)

DEALLOCATE (UUZ1)
```

```
DEALLOCATE (VVX1)

DEALLOCATE (VVY1)

DEALLOCATE (VVZ1)

DEALLOCATE (XXF1)

DEALLOCATE (YYF1)

DEALLOCATE (ZZF1)

DEALLOCATE (UUX2)

DEALLOCATE (UUY2)

DEALLOCATE (UUZ2)

DEALLOCATE (VVX2)

DEALLOCATE (VVY2)

DEALLOCATE (VVZ2)

DEALLOCATE (XXF2)

DEALLOCATE (YYF2)

DEALLOCATE (ZZF2)


ELSE

        WRITE (*,*) " ierror_1 = ", ierror_1

ENDIF


CLOSE (UNIT=11)

CLOSE (UNIT=10)


END PROGRAM
```

A.3.2 configuration_selector and distance_test subroutine

```
!------------------------------------------------------------------------------!

PROGRAM config_selector
```

```
!-------------------------------------------------------------------------------!


IMPLICIT NONE


LOGICAL :: accept_config


CHARACTER(LEN=5)  :: sym

CHARACTER(LEN=30) :: FMT


CHARACTER(LEN=5), ALLOCATABLE, DIMENSION(:) :: sym_1, sym_2


INTEGER, PARAMETER :: sgl = selected_real_kind(p=6)

INTEGER, PARAMETER :: dbl = selected_real_kind(p=12)


INTEGER :: ierror_1, ierror_2

INTEGER :: natms, natms_1, natms_2, mobile_mol

INTEGER :: config

INTEGER :: nconfig

INTEGER :: i, j, k


REAL(KIND=dbl) :: xx, yy, zz

REAL(KIND=dbl) :: rmax, rmin


REAL(KIND=dbl), ALLOCATABLE, DIMENSION(:) :: xx_1,  yy_1,  zz_1

REAL(KIND=dbl), ALLOCATABLE, DIMENSION(:) :: xx_2,  yy_2,  zz_2


FMT = "(A, 3F13.6)"


OPEN (UNIT=10, FILE='PARAM', STATUS='old', action='read', iostat=ierror_1)

IF (ierror_1 == 0) THEN

READ(10,*) nconfig

READ(10,*) natms_1
```

```
READ(10,*) natms_2

        READ(10,*) mobile_mol

READ(10,*) rmax

READ(10,*) rmin


        ALLOCATE(sym_1(natms_1))

ALLOCATE(xx_1(natms_1))

ALLOCATE(yy_1(natms_1))

ALLOCATE(zz_1(natms_1))

ALLOCATE(sym_2(natms_2*mobile_mol))

ALLOCATE(xx_2(natms_2*mobile_mol))

ALLOCATE(yy_2(natms_2*mobile_mol))

ALLOCATE(zz_2(natms_2*mobile_mol))


OPEN (UNIT=11, FILE='HISTORY', STATUS='old', action='READ', iostat=ierror_2)

IF (ierror_2 == 0) THEN

DO j = 1, nconfig

READ (11,*) natms

READ(11,10) config

10 FORMAT (14X, I10)

         k = 0

                        DO i = 1, natms

                                READ (11,*) sym, xx, yy, zz

                                IF (i <= natms_1) THEN

                                        sym_1(i) = sym

                                        xx_1(i)  = xx

                                        yy_1(i)  = yy

                                        zz_1(i)  = zz

                                ELSE

                                        k = k + 1

                                        sym_2(k) = sym

                                        xx_2(k)  = xx
```

```
                                        yy_2(k)  = yy

                                        zz_2(k)  = zz

                                ENDIF

ENDDO



CALL distance_test (natms_1, natms_2, mobile_mol, rmin, rmax, xx_1, yy_1, zz_1, xx_2, yy_2, zz_2, accept_config)


IF (accept_config) THEN

WRITE (*,*) natms

        WRITE (*,*) 'CONFIG ', config

                                k = 0

DO i = 1, natms

                                        IF (i <= natms_1) THEN

        WRITE (*, FMT) sym_1(i), xx_1(i), yy_1(i), zz_1(i)

        ELSE

        k = k + 1

                                                WRITE (*, FMT) sym_2(k), xx_2(k), yy_2(k), zz_2(k)

        ENDIF

                                ENDDO

        ENDIF

ENDDO

ELSE

WRITE(*,*) " ierror_2 = ", ierror_2

ENDIF


        DEALLOCATE (sym_1)

        DEALLOCATE (sym_2)

        DEALLOCATE (xx_1)

        DEALLOCATE (xx_2)

        DEALLOCATE (yy_1)

        DEALLOCATE (yy_2)
```

```
        DEALLOCATE (zz_1)
        DEALLOCATE (zz_2)

ELSE
WRITE(*,*) " ierror_1 = ", ierror_1
ENDIF

CLOSE (UNIT=10)

END PROGRAM config_selector



!-----------------------------------------------------------------------------------------------------------------------!
SUBROUTINE distance_test (natms_1, natms_2, mobile_mol, rmin, rmax, xx_1, yy_1, zz_1, xx_2, yy_2, zz_2, accept_config)
!-----------------------------------------------------------------------------------------------------------------------!
!c
!c                              SUBROUTINE TRANSLATOR
!c
!c     Version:  October 2014
!c
!c     Author:   Devis Di Tommaso, Etienne Gaines
!c

IMPLICIT NONE

integer, parameter :: sgl = selected_real_kind(p=6)
integer, parameter :: dbl = selected_real_kind(p=12)

!-----------------------------------------------!
! Declare calling parameter types & definitions !
!-----------------------------------------------!
! This is where we will have to define the number of variables
```

```
! The number of DMSO molecule will be important -> number or variables (?)

! This is where the min and max distance is defined

!-----------------------------------------------!


LOGICAL, INTENT(OUT) :: accept_config


INTEGER, INTENT(IN) :: natms_1, natms_2, mobile_mol


REAL(KIND=dbl), INTENT(IN)  :: rmax, rmin


REAL(KIND=dbl), INTENT(IN), DIMENSION(natms_1) :: xx_1, yy_1, zz_1

REAL(KIND=dbl), INTENT(IN), DIMENSION(natms_2*mobile_mol) :: xx_2, yy_2, zz_2


!-----------------------------------------------!

! Declare local variables types & definitions !

!-----------------------------------------------!

LOGICAL, DIMENSION(mobile_mol) :: accepted_molecule


INTEGER :: i, j, k, num_accepted


CHARACTER(LEN=30) :: FMT2, FMT3, FMT4


REAL(KIND=dbl) :: rxi, ryi, rzi

REAL(KIND=dbl) :: rxj, ryj, rzj

REAL(KIND=dbl) :: rxij, ryij, rzij

REAL(KIND=dbl) :: rijsq, rij


FMT2 = "(A5, ES10.3)"

FMT3 = "(A13, ES10.3)"

FMT4 = "(A6, ES10.3)"


num_accepted = 0
```

```
accepted_molecule = .FALSE.

accept_config = .FALSE.


DO i = 1, natms_1


        rxi = xx_1(i)

        ryi = yy_1(i)

        rzi = zz_1(i)


        DO k = 1, mobile_mol


               IF (accepted_molecule(k)) THEN

                  CYCLE

               ENDIF




               DO j = 1+natms_2*(k-1), natms_2*(k)




                        rxj = xx_2(j)

                        ryj = yy_2(j)

                        rzj = zz_2(j)


                        rxij = rxi - rxj

                        ryij = ryi - ryj

                        rzij = rzi - rzj


                        rijsq = rxij*rxij + ryij*ryij + rzij*rzij

                        rij = SQRT(rijsq)


                        IF ((rij > rmin) .AND. (rij < rmax)) THEN

                                accepted_molecule(k) = .true.
```

```
                        num_accepted = num_accepted + 1

                    ENDIF


                    IF (accepted_molecule(k)) EXIT


                ENDDO


            ENDDO


            IF (num_accepted == mobile_mol) THEN

                accept_config = .TRUE.

            ENDIF

            IF (accept_config) EXIT


    ENDDO


    END SUBROUTINE distance_test
```

Bibliography

[1] J. Bernstein, *Polymorphism in Molecular Crystals*, 2002.

[2] A. Nangia, *Accounts of Chemical Research*, 2008, **41**, 595–604.

[3] B. C. Hancock, E. Shalaev and S. L. Shamblin, *The Journal of pharmacy and pharmacology*, 2002, **54**, 1151–1152.

[4] F. Wöhler and J. Liebig, *Ann. Pharm.*, 1832, 249 282.

[5] J. Thun, L. Seyfarth, J. Senker, R. E. Dinnebier and J. Breu, *Angewandte Chemie - International Edition*, 2007, **46**, 6729–6731.

[6] G. M. Day, A. V. Trask, W. D. Motherwell and W. Jones, *Chemical Communications*, 2006, **1**, 54–56.

[7] W. Bragg, *Nature*, 1939, **143**, 73–74.

[8] M. Shahat, *Acta Crystallographica*, 1952, 763–768.

[9] M. N. G. James and G. J. B. Williams, *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, 1974, **30**, 1249–1257.

[10] A. V. Trask, D. A. Haynes, W. D. Motherwell and W. Jones, *Chemical Communications*, 2006, 51–53.

[11] A. Biswas, L. K. Das and A. Ghosh, *Polyhedron*, 2013, **61**, 253–261.

[12] A. J. Cruz-Cabeza, S. M. Reutzel-Edens and J. Bernstein, *Chem. Soc. Rev.*, 2015, **44**, 8619–8635.

[13] L. Yu, G. A. Stephenson, C. A. Mitchell, C. A. Bunnell, S. V. Snorek, J. J. Bowyer, T. B. Borchardt, J. G. Stowell and S. R. Byrn, *Journal of the American Chemical Society*, 2000, **122**, 585–591.

[14] W. C. McCrone, *Physics and Chemistry of the Organic Solid State*, Interscience Publishers, New York, 1965, pp. 725–767.

[15] M. K. Singh, A. Banerjee and P. K. Gupta, *Crystal Growth and Design*, 2012, **12**, 732–741.

[16] T. Bartels-Rausch, *Nature*, 2013, **494**, 27–29.

[17] B. J. Murray, T. W. Wilson, S. Dobbie, Z. Cui, S. M. Al-Jumur, O. Möhler, M. Schnaiter, R. Wagner, S. Benz, M. Niemand, H. Saathoff, V. Ebert, S. Wagner and B. Kärcher, *Nature Geoscience*, 2010, **3**, 233–237.

[18] E. G. Hammerschmidt, *Industrial and Engineering Chemistry Research*, 1934, **26**, 851–855.

[19] E. D. Sloan, *Nature*, 2003, **426**, 353–359.

[20] R. Velázquez-Castillo, J. Reyes-Gasga, D. I. García-Gutierrez and M. Jose-Yacaman, *Journal of Materials Research*, 2006, **21**, 1484–1489.

[21] J. D. Harper, C. M. Lieber and P. T. Lansbury Jr., *Chemistry & Biology*, 1997, **4**, 951–959.

[22] D. M. Walsh, A. Lomakin, G. B. Benedek, M. M. Condron and D. B. Teplow, 1997, **272**, 22364–22372.

[23] S. Piana, F. Jones and J. D. Gale, *Journal of the American Chemical Society*, 2006, **128**, 13568–13574.

[24] K. Hachiya, M. Sasaki, T. Ikeda, N. Mikami and T. Yasunaga, *Journal of Physical Chemistry*, 1984, **88**, 27–31.

[25] A. E. Nielsen, *Journal of Crystal Growth*, 1984, **67**, 289–310.

[26] A. Nielsen and J. Toft, *Journal of Crystal Growth*, 1984, **67**, 278–288.

[27] O. S. Pokrovsky and J. Schott, *Environmental Science and Technology*, 2002, **36**, 426–432.

[28] I. Sethmann, A. Putnis, O. Grassmann and P. Löbmann, *American Mineralogist*, 2005, **90**, 1213–1217.

[29] J. Xu, C. Yan, F. Zhang, H. Konishi, H. Xu and H. H. Teng, *Proceedings of the National Academy of Sciences*, 2013, **110**, 17750–17755.

[30] D. D. Tommaso and N. H. De Leeuw, *Physical Chemistry Chemical Physics*, 2010, **12**, 894–901.

[31] D. Mealey, D. M. Croker and . C. Rasmuson, *CrystEngComm*, 2015, **17**, 3961–3973.

[32] D. Khamar, J. Zeglinski, D. Mealey and A. C. Rasmuson, *Journal of the American Chemical Society*, 2014, **136**, 11664–11673.

[33] D. Mealey, J. Zeglinski, D. Khamar and . C. Rasmuson, *Faraday Discussions*, 2015, **179**, 309–328.

[34] J. J. De Yoreo, *Reviews in Mineralogy and Geochemistry*, 2003, **54**, 57–93.

[35] P. Hohenberg and B. Halperin, *Reviews of Modern Physics*, 1977, **49**, 435–479.

[36] P. Chaikin and T. Lubensky, *Principles of condensed matter physics*, Cambridge University Press, Cambridge, 1995.

[37] D. Kashchiev, *Nucleation: Basic Theory with Applications.*, Heinemann, Oxford, Butterwort edn, 1999.

[38] D. Zahn, *ChemPhysChem*, 2015, **16**, 2069–2075.

[39] R. J. Davey, S. L. M. Schroeder and J. H. Ter Horst, *Angewandte Chemie - International Edition*, 2013, **52**, 2167–2179.

[40] I. Weissbuch, M. Lahav and L. Leiserowitz, *Crystal Growth and Design*, 2003, **3**, 125–150.

[41] P. G. Vekilov, *Crystal Growth and Design*, 2010, **10**, 5007–5019.

[42] H. Henry Teng, *Elements*, 2013, **9**, 189–194.

[43] F. C. Meldrum and R. P. Sear, *Science (New York, N.Y.)*, 2008, **322**, 1802–1803.

[44] A. S. Myerson and B. L. Trout, *Science*, 2013, **341**, 855–856.

[45] W. C. M. Lewis, *Chem. Rev.*, 1931, **8**, 81–165.

[46] R. M. Ginde and A. S. Myerson, *Journal of Crystal Growth*, 1992, **116**, 41–47.

[47] D. Erdemir, S. Chattopadhyay, L. Guo, J. Ilavsky, H. Amenitsch, C. U. Segre and A. S. Myerson, *Physical Review Letters*, 2007, **99**, 1–4.

[48] J. Huang, T. C. Stringfellow and L. Yu, *Journal of the American Chemical Society*, 2008, **130**, 13973–13980.

[49] A. S. Myerson and P. Y. Lo, *Journal of Crystal Growth*, 1991, **110**, 26–33.

[50] Y. C. Chang and A. S. Myerson, *AIChE Journal*, 1986, **32**, 1567–1569.

[51] C. K. Meng and J. B. Fenn, *Organic Mass Spectrometry*, 1991, **26**, 542–549.

[52] Z. Takats, S. C. Nanita, R. G. Cooks, G. Schlosser and K. Vekey, *Analytical Chemistry*, 2003, **75**, 1514–1523.

[53] P. Nemes, G. Schlosser and K. Vékey, *Journal of Mass Spectrometry*, 2005, **40**, 43–49.

[54] D. Zhang, L. Wu, K. Koch and R. Cooks, *European Journal of Mass Spectrometry*, 1999, **5**, 353.

[55] N. Toyama, J. Y. Kohno, F. Mafuné and T. Kondow, *Chemical Physics Letters*, 2006, **419**, 369–373.

[56] J. R. Bourne and R. J. Davey, *Journal of Crystal Growth*, 1976, **36**, 287–296.

[57] M. K. Singh and A. Banerjee, *CrystEngComm*, 2013, **15**, 4143.

[58] K. Srinivasan and J. N. Sherwood, *Crystal Growth and Design*, 2005, **5**, 1359–1370.

[59] K. Srinivasan and J. N. Sherwood, *Crystal Growth and Design*, 2011, **11**, 5010–5018.

[60] B. Akn, M. Öner, Y. Bayram and K. D. Demadis, *Cryst. Growth Des.*, 2008, **8**, 1997–2005.

[61] C. O. Perera, I. C. Hallett, T. T. Nguyen and J. C. Charles, *Journal of Food Science*, 1990, **55**, 1066–1069.

[62] G. L. R. G. C. J. Horenstein, B. K.; Hernandez, *Wat. Sci. Technol.*, 1990, **22**, 183191.

[63] L. S. Parvaneh, D. Donadio and M. Sulpizi, *The Journal of Physical Chemistry C*, 2016, **120**, 4410–4417.

[64] A. Verch, A. S. Côté, R. Darkins, Y. Y. Kim, R. Van De Locht, F. C. Meldrum, D. M. Duffy and R. Kröger, *Small*, 2014, **10**, 2697–2702.

[65] B. Cantaert, E. Beniash and F. C. Meldrum, *Chemistry - A European Journal*, 2013, **19**, 14918–14924.

[66] Q. Jiang and M. D. Ward, *Chemical Society Reviews*, 2014, **43**, 2066–2079.

[67] E. Loste, R. J. Park, J. Warren and F. C. Meldrum, *Advanced Functional Materials*, 2004, **14**, 1211–1220.

[68] F. C. Meldrum and H. Cölfen, *Chemical Reviews*, 2008, **108**, 4332–4432.

[69] A. S. Schenk, E. J. Albarracin, Y. Y. Kim, J. Ihli and F. C. Meldrum, *Chemical Communications*, 2014, **50**, 4729–4732.

[70] Y. W. Wang, H. K. Christenson and F. C. Meldrum, *Chemistry of Materials*, 2014, **26**, 5830–5838.

[71] M. Lahav and L. Leiserowitz, *Chemical Engineering Science*, 2001, **56**, 2245–2253.

[72] R. J. Davey, N. Blagden, G. D. Potts and R. Docherty, 1997, **7863**, 1767–1772.

[73] J. Chen and B. L. Trout, *Journal of Physical Chemistry B*, 2010, **114**, 13764–13772.

[74] R. J. Davey, G. Dent, R. K. Mughal and S. Parveen, *Crystal Growth and Design*, 2006, **6**, 1788–1796.

[75] S. Parveen, R. J. Davey, G. Dent and R. G. Pritchard, *Chemical Communications*, 2005, **12**, 1531–1533.

[76] J. Chen and B. L. Trout, *J. Phys. Chem. B*, 2008, **1**, 7794–7802.

[77] S. Chattopadhyay, D. Erdemir, J. M. B. Evans, J. Ilavsky, H. Amenitsch, C. U. Segre and A. S. Myerson, *Crystal Growth and Design*, 2005, **5**, 523–527.

[78] R. J. Davey, N. Blagden, S. Righini, H. Alison, M. J. Quayle and S. Fuller, *Crystal Growth & Design*, 2001, **1**, 59–65.

[79] R. J. Davey, G. Dent, R. K. Mughal and S. Parveen, *Crystal Growth and Design*, 2006, **6**, 1788–1796.

[80] C. S. Towler and L. S. Taylor, *Crystal Growth and Design*, 2007, **7**, 633–638.

[81] M. Svard and . C. Rasmuson, *Crystal Growth and Design*, 2013, **13**, 1140–1152.

[82] D. Di Tommaso, *CrystEngComm*, 2013, **15**, 6564.

[83] B. Malmberg and A. C. Rasmuson, *Crystal Growth and Design*, 2012, **12**, 4340–4348.

[84] S. S. L. Price, *Accounts of Chemical Research*, 2009, **42**, 117–126.

[85] S. L. Price, M. Leslie, G. W. A. Welch, M. Habgood, L. S. Price, P. G. Karamertzanisc and G. M. Day, *Phisical Chemistry Chemical Physics*, 2010, 8478–8490.

[86] A. Gavezzotti, *Chem. Eur. J.*, 1999, **5**, 567–576.

[87] S. Hamad, C. E. Hughes, C. R. a. Catlow and K. D. M. Harris, *J. Phys. Chem. B*, 2008, **112**, 7280–7288.

[88] J. Kästner, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, **1**, 932–942.

[89] C. Abrams and G. Bussi, *Entropy*, 2014, **16**, 163–199.

[90] M. Salvalaglio, T. Vetter, F. Giberti, M. Mazzotti and M. Parrinello, *Journal of American Chemical Society*, 2012, **134**, 17221–17233.

[91] T. C. Lewis, D. A. Tocher, G. M. Day and S. L. Price, *CrystEngComm*, 2003, **5**, 3–9.

[92] M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti and M. Parrinello, *Proceedings of the National Academy of Sciences*, 2015, **112**, 6–14.

[93] R. J. Davey, N. Blagden, S. Righini, H. Alison and E. S. Ferrari, *Journal of Physical Chemistry B*, 2002, **106**, 1954–1959.

[94] M. A. Palafox, M. Gill and J. L. Núñez, *Spectroscopy Letters*, 1996, **29**, 609–629.

[95] P. A. Williams, C. E. Hughes, G. K. Lim, B. M. Kariuki and K. D. M. Harris, *Crystal Growth and Design*, 2012, **12**, 3104–3113.

[96] P. A. Williams, C. E. Hughes, G. K. Lim, B. M. Kariuki and K. D. M. Harris, *Crystal Growth and Design*, 2012, **12**, 3104–3113.

[97] W. D. Kumler, *Journal of Organic Chemistry*, 1955, **20**, 700–706.

[98] N. Bjerrum, *Zeitschrift für Physikalische Chemie*, 1923, **104U**, 147.

[99] M. Svärd, F. L. Nordström, T. Jasnobulka and A. C. Rasmuson, *Crystal Growth and Design*, 2010, **10**, 195–204.

[100] S. Gracin and . C. Rasmuson, *Crystal Growth and Design*, 2004, **4**, 1013–1023.

[101] E. Simone, G. Steele and Z. K. Nagy, *CrystEngComm*, 2015, **17**, 9370–9379.

[102] C. J. Brown and M. Ehrenberg, *Acta Crystallographica Section C*, 1985, **41**, 30441–30443.

[103] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka, *Journal of Chemical Physics*, 2005, **123**, year.

[104] S. Grimme, *Journal of computational chemistry*, 2006, **27**, 1787–1799.

[105] R. A. Sullivan, R. J. Davey, G. Sadiq, G. Dent, K. R. Back, J. H. Ter Horst, D. Toroz and R. B. Hammond, *Crystal Growth and Design*, 2014, **14**, 2689–2696.

[106] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *Journal of Physical Chemistry B*, 2009, **113**, 6378–6396.

[107] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *Journal of Physical Chemistry B*, 2009, **113**, 4538–4543.

[108] R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, *Physical Chemistry Chemical Physics*, 2011, **13**, 10908–10922.

[109] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Computer Physics Communications*, 2014, **185**, 604–613.

[110] L. A. Montero, *Generacion de Celdas con Agregados Moleculares Optimizados [Online].*, 1998, `http://karin.fq.uh.cu/mmh/`.

[111] L. A. Montero, A. M. Esteva, J. Molina, A. Zapardiel, L. Hernández, H. Márquez and A. Acosta, *Journal of the American Chemical Society*, 1998, **120**, 12023–12033.

[112] E. Tang, D. Di Tommaso and N. H. de Leeuw, *Physical Chemistry Chemical Physics*, 2010, **12**, 13804.

[113] A. Théorêt, *Spectrochimica Acta Part A: Molecular Spectroscopy*, 1971, **27**, 11–18.

[114] J. ter Horst, R. Geertman and G. van Rosmalen, *Journal of Crystal Growth*, 2001, **230**, 277–284.

[115] N. Blagden and R. J. Davey, *Crystal Growth and Design*, 2003, **3**, 873–885.

[116] D. Musumeci, C. A. Hunter and J. F. McCabe, *Crystal Growth and Design*, 2010, **10**, 1661–1664.

[117] C. E. Hughes, P. A. Williams and K. D. Harris, *Angewandte Chemie - International Edition*, 2014, **53**, 8939–8943.

[118] S. Goedecker, M. Teter and J. Hutter, *Physical Review B*, 1996, **54**, 1703–1710.

[119] Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2008, **120**, 215–241.

[120] B. J. Lynch, Y. Zhao and D. G. Truhlar, *Journal of Physical Chemistry A*, 2003, **107**, 1384–1388.

[121] I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory and Computation*, 2010, **6**, 2872–2887.

[122] M. P. Gaigeot and M. Sprik, *The Journal of Physical Chemistry B*, 2004, **108**, 7458–7467.

[123] E. Tang, D. Di Tommaso and N. H. de Leeuw, *The Journal of Chemical Physics*, 2009, **130**, 234502.

[124] P. C. Sadek, *Laminar flow*, Hoboken, Wiley edn, 2004.

[125] J. R. Pliego and J. M. Riveros, *Journal of Physical Chemistry A*, 2001, **105**, 7241–7247.

[126] D. Di Tommaso and K. L. Watson, *The journal of physical chemistry. A*, 2014, **118**, 11098–113.

[127] P. Ectors, P. Duchstein and D. Zahn, *Crystal Growth and Design*, 2014, **14**, 2972–2976.

[128] H. G. Brittain and D. J. W. Grant, *Polymorphism in Pharmaceutical Solids*, New York, edn, 1999, pp. 279–330.

[129] *ANDAs: Pharmaceutical Solid Polymorphism Chemistry, Manufacturing, and Controls Information*, U.s. department of health and human services food and drug administration center for drug evaluation and research technical report, 2007.

[130] D. Braga, F. Grepioni, M. Lucia and P. Marco, *Structure and Bonding*, 2009, **132**, 25–50.

[131] C. Stoica, P. Verwer, H. Meekes, P. J. C. M. van Hoof, F. M. Kaspersen and E. Vlieg, *Crystal Growth & Design*, 2004, **4**, 765–768.

[132] A. Spitaleri, C. A. Hunter, J. F. McCabe, M. J. Packer and S. L. Cockroft, *CrystEngComm*, 2004, **6**, 489–493.

[133] R. J. Davey, K. R. Back and R. A. Sullivan, *Faraday Discussions*, 2015, **179**, 9–26.

[134] E. Ruiz-Agudo and C. V. Putnis, *Mineralogical Magazine*, 2012, **76**, 227–253.

[135] I. Weissbuch, L. Addadi, M. Lahav and L. Leiserowitz, *Science*, 1991, **253**, 637–645.

[136] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. De Jong, *Computer Physics Communications*, 2010, **181**, 1477–1489.

[137] M. J. Frisch, G. W. Trucks, H. Schlegel, G. E. Scuseria, M. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, L. J. Sonnenberg, M. Hada, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpa, J. Cioslowski and D. J. Fox, *Inc., Wallingford CT*, 2009, **121**, 150–166.

[138] Y. Zhao and D. G. Truhlar, *Journal of Chemical Theory and Computation*, 2008, **4**, 1849–1868.

[139] R. Peverati and D. G. Truhlar, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2014, **372**, 20120476–20120476.

[140] E. R. Johnson, I. D. Mackie and G. A. DiLabio, *Journal of Physical Organic Chemistry*, 2009, **22**, 1127–1135.

[141] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. a. Kollman and D. a. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[142] H. J. Berendsen, J. R. Grigera and T. P. Straatsma, *Journal of Physical Chemistry*, 1987, **91**, 6269–6271.

[143] C. I. Bayly, P. Cieplak, W. D. Cornell and P. a. Kollman, *The Journal of Physical Chemistry*, 1993, **97**, 10269–10280.

[144] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, *Physical chemistry chemical physics : PCCP*, 2010, **12**, 7821–39.

[145] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *Journal of the American Chemical Society*, 1995, **117**, 5179–5197.

[146] T. E. Cheatham, P. Cieplak and P. A. Kollman, *Journal of Biomolecular Structure and Dynamics*, 1999, **16**, 845–862.

[147] J. Wang, P. Cieplak and P. A. Kollman, *Journal of Computational Chemistry*, 2000, **21**, 1049–1074.

[148] V. Hornak, R. Abel, A. Okuyr, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712–725.

[149] B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435–447.

[150] D. van der Spoel, *GROMACS [Online]*, 2015, `http://www.gromacs.org/`.

[151] A. Barducci, G. Bussi and M. Parrinello, *Physical Review Letters*, 2008, **100**, 1–4.

[152] A. Laio and F. L. Gervasio, *Reports on Progress in Physics*, 2008, **71**, 1–22.

[153] P. Friant-Michel and M. F. Ruiz-López, *ChemPhysChem*, 2010, **11**, 3499–3504.

[154] T. Krishnan, W. C. Duer, S. Goldman and J.-L. Fortier, *Canadian Journal of Chemistry*, 1979, **57**, 530–537.

[155] X. Biarnés, F. Pietrucci, F. Marinelli and A. Laio, *Computer Physics Communications*, 2012, **183**, 203–211.

[156] S. S. Kumar and A. Nangia, *Crystal Growth and Design*, 2014, **14**, 1865–1881.

[157] E. Gaines, K. Maisuria and D. Di Tommaso, *CrystEngComm*, 2016, **18**, 2937–2948.

[158] E. J. Cohn, J. T. Edsall, J. G. Kirkwood, H. Mueller, J. L. Oncley and G. Scatchard, *The American Journal of the Medical Sciences*, 1943, **205**, 884.

[159] J. Ho, A. Klamt and M. L. Coote, *Journal of Physical Chemistry A*, 2010, **114**, 13442–13444.

[160] H. Do and N. A. Besley, *Journal of Chemical Physics*, 2012, **137**, year.

[161] E. Sánchez-García, M. Studentkowski, L. A. Montero and W. Sander, *ChemPhysChem*, 2005, **6**, 618–624.

[162] M. Salvalaglio, F. Giberti and M. Parrinello, *Acta Crystallographica Section C: Structural Chemistry*, 2014, **70**, 132–136.

[163] D. Toroz, R. B. Hammond, K. J. Roberts, S. Harris and T. Ridley, *Journal of Crystal Growth*, 2014, **401**, 38–43.

[164] W. Du, A. J. Cruz-Cabeza, S. Woutersen, R. J. Davey and Q. Yin, *Chemical Science*, 2015, **6**, 3515–3524.

[165] I. Weissbuch, V. Y. Torbeev, L. Leiserowitz and M. Lahav, *Angewandte Chemie - International Edition*, 2005, **44**, 3226–3229.

[166] M. N. Bhat and Dharmaprakash, *Journal of Crystal Growth*, 2002, **242**, 245–252.

[167] A. F. Wallace, L. O. Hedges, A. Fernandez-martinez, P. Raiteri, J. D. Gale, G. A. Waychunas, S. Whitelam, J. F. Banfield and J. J. De, 2013, **341**, 885–889.

[168] J. Anwar and D. Zahn, *Angewandte Chemie - International Edition*, 2011, **50**, 1996–2013.

[169] P. Raiteri, R. Demichelis, J. D. Gale, M. Kellermeier, D. Gebauer, D. Quigley, L. B. Wright and T. R. Walsh, *Faraday Discussions*, 2012, **159**, 61–85.

[170] E. Gaines and D. Di Tommaso, *Pharmaceutics*, 2018, **10**, 1–15.

[171] D. A. J. M. Voogd. J, Verzijl. B. H. M, *Acta Crystallographica Section B*, 1980, **36**, 2805–2806.

[172] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson and M. C. Payne, *Zeitschrift für Kristallographie*, 2005, **220**, 567–570.

[173] J. Hutter, M. Iannuzzi, F. Schiffmann and J. Vandevondele, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2014, **4**, 15–25.

[174] J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865–3868.

[175] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *Journal of Chemical Physics*, 2010, **132**, 1–19.

[176] C. J. Cramer, *Essentials of Computational Chemistry Theories and Models*, 2004, vol. 42, pp. 334–342.

[177] D. Toroz, I. Rosbottom, T. D. Turner, D. M. Corzo, R. B. Hammond, X. Lai and K. J. Roberts, *Faraday Discussions*, 2015, **179**, 79–114.

[178] T. S. Hofer, H. T. Tran, C. F. Schwenk and B. M. Rode, *Journal of Computational Chemistry*, 2004, **25**, 211–217.

[179] M. Wolthers, D. Di Tommaso, Z. Du and N. H. De Leeuw, *CrystEngComm*, 2013, **15**, 5506–5514.

[180] E. Ruiz-Agudo, D. Di Tommaso, C. V. Putnis, N. H. De Leeuw and A. Putnis, *Crystal Growth and Design*, 2010, **10**, 3022–3035.

[181] D. Di Tommaso, E. Ruiz-Agudo, N. H. De Leeuw, A. Putnis and C. V. Putnis, *Physical Chemistry Chemical Physics*, 2014, **16**, 7772–7785.

[182] S. Kerisit and S. C. Parker, *Chemical communications (Cambridge, England)*, 2004, **1**, 52–53.

[183] M. Wolthers, D. Di Tommaso, Z. Du and N. H. De Leeuw, *Physical Chemistry Chemical Physics*, 2012, **14**, 15145–15157.

[184] F. Bertaut, *Compt. Rendu.*, 1958, **246**, 3447.

[185] P. W. Tasker, *Journal of Physics C: Solid State Physics*, 1979, **12**, 4977–4984.

[186] B. M. Rode, C. F. Schwenk, T. S. Hofer and B. R. Randolf, *Coordination Chemistry Reviews*, 2005, **249**, 2993–3006.

Contributions

List of publications

- Gaines E., Maisuria K. and Di Tommaso D., The role of solvent in the self-assembly of *m*-aminobenzoic acid: a density functional theory and molecular dynamics study, ***CrystEngComm***, 2016, **18**, 2937-2948

- Gaines E. and Di Tommaso D., Solvation and Aggregation of Meta-Aminobenzoic Acid in Water: Density Functional Theory and Molecular Dynamics Study. *Pharmaceutics* **2018**, *10*, 12

Oral interventions and poster presentations

- BACG2015 - 46th Annual Conference of the "British Association for Crystal Growth" (Queen Mary - 21^{st} to 23^{rd} of June 2015) - **Poster presentation**, 2^{nd} prize.

- Thomas Young Centre - 10^{th} Anniversary Symposium (London, 17^{th} to 19^{th} of February 2016) - **Poster presentation**

- Postgraduate research symposium (Queen Mary, 22^{nd} of February 2016) - **Poster presentation**

- BACG2016 - 47^{th} Annual Conference of the "British Association for Crystal Growth" (University of Leeds - 27^{th} to 29^{th} of June 2016) - **Oral presentation**

- GMDM4 - The "Granada-Münster Discussion Meeting" (University of Münster - 24^{th} and 25^{th} of November 2016) - **Oral presentation**

- Postgraduate research symposium (Queen Mary, 20^{th} of February 2017) - **Oral presentation**

- Modelling of Advanced Functional Materials using Terascale Computing (Cardiff University - Wednesday 6^{th} to Friday 8^{th}, April 2016) - **Oral presentation**

- "Recent Appointees in Materials Modelling" (London, 23^{rd} of March 2016) - **Poster presentation**

- BACG2017 - 48^{th} Annual Conference of the "British Association for Crystal Growth" (University of Manchester - 27^{th} to 30^{th} of June 2016) - **Oral presentation**